CERTIFICATE OF MAILING BY "EXPRESS MAIL"
Express Mail Label No. EL615209879US Date of Deposit: November 2, 2001
I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 C.F.R. § 1.10 on the date indicated above and is addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231. Tammy James

APPLICATION IN

THE UNITED STATES PATENT AND TRADEMARK OFFICE

FOR

ADAPTER-DIRECTED DISPLAY SYSTEMS

INVENTORS:

Caili Wang
95 Judah St., Apt. 1 San Francisco, CA, 94122
Citizenship: Peoples' Republic of China

Pingyu Zhong
1475 Tyler Parkway, Mountain View, CA 94040
Citizenship: USA

Xinwei Wang
7221 Via Vico, San Jose, CA 95129
Citizenship: Peoples' Republic of China

HOWREY SIMON ARNOLD & WHITE, LLP
750 Bering Drive, Houston, Texas 77057
Attorney Docket No. 13403.0005NPUS00

Express Mail Post Office

ADAPTER-DIRECTED DISPLAY SYSTEM

TECHNICAL FIELD

1 This invention is in the field of display technology. Specifically, the invention relates to the generation of adapter-directed display systems for exogenous display of polypeptides on genetic packages. The compositions and methods embodied in the present invention are particularly useful for identifying from a vast repertoire of polypeptides those individual members exhibiting desired properties.

BACKGROUND OF THE INVE NTION

2 The display of polypeptides on the surface of genetic packages represents a powerful methodology for carrying out molecule evolution in the laboratory. The ability to construct libraries of enormous molecular diversity and to select for molecules with desired properties has made this technology applicable to a wide range of problems. The origins of phage display date to the mid-1980s when George Smith first expressed an exogenous segment of a protein on the surface of bacteriophage M13 virus particles by fusing the exogenous sequence to a phage coat protein (*Science* (1985) **228**: 1315-1317). Two groundbreaking concepts emerged from Smith's initial experiment. First, the experiment suggested that a vast diverse repertoire of polypeptides could be constructed in which individual phage particles display unique polypeptides. Second, the experiment confirmed a direct physical link between phenotype and genotype. That is, the phage displaying the desired polypeptide also harbors the DNA encoding the polypeptide, which can be readily isolated for subsequent analyses. McCafferty and Ladner extended these concepts to screen repertoire of polypeptides such as single-chain antibodies displayed on the surface of phage particles (U.S. Patent Nos. 5,969,108 and

5,837,500). Since then, phage display has become a popular technique for protein engineering.

3 A range of display systems have been developed based on George Smith's findings. These systems can be broadly classified into two categories. The first generation system is a one-vector system. The vector in this system contains the entire phage genome, insert therein an exogenous sequence in-frame with a coat protein gene. Because the resulting phage particles carry the entire phage genomes, they are relatively unstable and less infectious. The second generation system, commonly referred to as the phagemid system, has two components: (1) a phagemid vector carrying the exogenous sequence fused to phage coat protein, and a phage-derived origin of replication to allow packaging the phagemid into a phage particle; and (2) a helper phage vector carrying all other sequences required for phage packaging. The helper vector is typically replication-defective such as M13KO7 helper vector manufactured by Amersham Pharmacia Biotech and its derivative VCSM13 that is produced by Stratagen. Upon superinfection of a bacterial cell with the helper phages, newly packaged phages carrying the phagemid vector and displaying the exogenous sequence are produced.

4 As such, the prior phagemid system requires fusion of the exogenous sequence to at least part of a phage outer-surface sequence (i.e. the coat sequence). The fusion or display sites most commonly used are within genes III and VIII of M13 bacteriophage, although genes VI, VII and IX fusions have been reported. However, these fusion systems bear a number of pronounced limitations. First and foremost, the expression of coat proteins is toxic to the host cells, thus tight regulation of the coat-fusion must be monitored. Even so, the unavoidable promoter leakage can cause loss of members of a diverse library. Maintaining the stability of a library is especially critical for generating a vast diverse repertoire of molecules (such as antigen-binding units) with a complexity

of at least 10^9. Second, expression of certain coat proteins such as the gene III product (pIII) can render host cells resistant to infection with helper phage required for the production of progeny phage particles. Third, the fusion format including the gene III and gene VIII phage display systems restrict the point of insertion to the 5' end of the outer-surface sequence. The exogenous polypeptide thus must be linked to the N-terminus of the outer-surface proteins. Consequently, cDNA libraries containing fragments of coding sequences of all reading frames cannot be fully displayed by these fusion systems due to frequent disruption of reading frames by internal stop codons. Furthermore, the fusion system is unstable due to recombination between the fusion and the wildtype outer-surface protein that is typically provided by a helper vector. Finally, since the phagemid vector contains at least a portion of the outer-surface sequences, large exogenous sequence may not be efficiently expressed because of low transformation efficiency of a large vector. Transformation efficiency, however, is a critical factor for the production of libraries of high complexity.

Various modifications to the fusion phagemid system have been described. WO 91/17271 proposes construction of a phage display system in which the exogenous sequence is displayed via interaction of a "tag" and a "tag ligand." The system contains a phage genomic vector that carries the exogenous sequence joined to a tag sequence. The same vector carries a tag ligand sequence fused in-frame with a coat protein gene. Upon infection of a host cell with the vector, it is speculated that phage particles expressing the exogenous sequences would be produced. However, the disclosure of WO 91/17271 does not provide a teaching which enables the general idea to be carried out. For example, WO 91/17271 does not demonstrate that any sequence has been displayed on the surface of phage particle via the interaction between a "tag" and a "ligand;" nor has it demonstrated that the protein, if expressed, retains biological activity. Furthermore, because the proposed system employs a phage genomic vector

carrying all phage coat protein genes in the same vector, the system inevitably inherits all limitations and drawbacks as described above.

6 Crameri et al. devised a system to display cDNA products, in which *Fos* oncogene was inserted adjacent to the exogenous sequence to be displayed on a phagemid vector, and *Jun* oncogene was inserted adjacent to gene III on the same vector (see Crameri et al. (1993) *Gene* **137**:69-75). These two fusion sequences were placed under the control of two separate promoters. The Crameri approach exploits the preferential interaction between fos and jun proteins: as the Fos-exogenous polypeptide is expressed and secreted into the periplasmic space, it forms a complex with pIII-Jun which is then packaged into the phage particles upon superinfection with M13KO7 helper phage. Although the exogenous sequence in this system is not directly linked to an outer-surface sequence, the constitutive expression of phage coat protein pIII under a separate promoter of the same vector still causes substantial toxicity to the host cells.

7 Another variant similar to the Crameri system is the "cysteine-coupled" display system described in WO 01/05950. The attachment and display of the exogenous polypeptide are mediated by the formation of disulfide bond between two cysteine residues, one of which is contained in the exogenous sequence, and the other is inserted in the outer-surface sequence. The one vector system described in WO 01/05950 is a phagemid vector carrying two separate promoter-controlled expression cassettes: one expresses the exogenous sequence, and the another expresses the coat protein pIII. The two-vector system described in WO 01/05950 contains a phagemid vector carrying an exogenous sequence, and a plasmid expressing the coat protein pIII. The two vectors are used to co-transfect E. Coli cells. Upon superinfection with the helper phages, M13KO7, the phagemid and/or the plasmid are packaged into the resulting phage particles. Although this system avoids the expression of a fusion comprising the exogenous protein linked to an outer-surface protein, the system again fails to minimize

the toxicity of coat proteins to the host cells because of the constitutive expression of the coat protein pIII in either the one-vector or the two-vector system. Furthermore, the two-vector system described in WO 01/05950 inevitably produces phage particles with mispackaged vectors carrying the outer-surface sequences and not the exogenous gene upon infection of the helper phages. Mispackaging is a well-known problem associated with two-vector system. It has been shown that the pIII-supplementing plasmid vectors were mispackaged into helper phage particles (Rondot et al. (2000) *Nature* **19**: 75-78).

8 Finally, the aforementioned prior phage display systems are not compatible with other display systems, such as a bacterial display system. To present the same phage-displayed exogenous sequence directly onto a bacterial cell, the exogenous sequence must first be subcloned into a bacterial display vector.

9 Thus, there remains a considerable need for improved compositions and methods for exogenous display on genetic packages. An ideal system would avoid the drawbacks of the previously reported systems. The present invention satisfies these needs and provides related advantages as well.

SUMMARY OF THE INVENTION

10 A principal aspect of the present invention is the design of systems that enable display of polypeptides not linked to any outer-surface sequences of a genetic package via peptide bonds. The experimental design provides an unprecedented flexibility for the presentation and/or selection of proteins with desired properties on a genetic package such as a phage particle. The technical advantages of the subject phage-display system are manifold. First, the system avoids all drawbacks associated with expression of the outer-surface proteins by the expression vectors. As mentioned above, the drawbacks include (1) high toxicity to the host cell as a result of constitutive expression

of the outer-surface sequences; (2) resistance of host cells to the infection of helper phages that is required for the production of progeny phage particles; (3) limitation on the orientation of the proteins to be displayed because of the unidirectional display of N-terminal fusion product; and (4) instability of the fusion product due to recombination between the fusion outer-surface sequence and the wildtype outer-surface sequence which is typically provided by the helper vector. Second, the system eliminates the possibility of mispackaging plasmids carrying the outer-surface sequences and not the gene of interest; such plasmids are used in the two-vector system described in WO 01/0595. While avoiding these and other intrinsic shortcomings of the prior display systems, the subject system further provides the flexibility of presenting one copy (monovalent display) or multiple copies (multivalent display) of a polypeptide per genetic package. The subject systems are particularly useful for expressing and screening a vast diverse repertoire of polypeptides (i.e. antigen-binding units) based on their ability to bind molecules of particular interest. The polypeptides displayed by the subject systems are functional.

Accordingly, the present invention provides an adapter-directed display system for displaying an exogenous polypeptide on the outer surface of a genetic package. The system comprises: (a) an expression vector comprising a coding sequence that encodes the exogenous polypeptide fused in-frame to a first adapter sequence, wherein the vector is devoid of outer-surface sequences encoding any functional outer-surface proteins of the genetic package; (b) a helper vector comprising outer-surface sequences encoding outer-surface proteins necessary for packaging the genetic package, wherein at least one of the outer-surface protein is fused in-frame to a second adapter, said first and second adapter acting, when the polypeptide is produced in a suitable host cell, to cause the display of the polypeptide via pairwise interaction between the first and second adapters.

12 The employed genetic package can be viruses, cells, and spores. In one embodiment, the subject system is a phage display system. The outer-surface sequences encode functional coat proteins of a phage. The preferred outer-surface sequences encode functional coat proteins of a phage such as a filamentous phage. Preferred outer-surface sequences are selected from the group consisting of gene III, gene VI, gene VII, gene VIII, and gene IX of a filamentous phage.

13 In another embodiment the subject system is a bacterial display system. The outer-surface sequences excluded in the expression vector but included in the bacterial helper vector encode bacterial outer-surface proteins. Preferred outer-surface proteins are selected from the group consisting of Lpp-OmpA, TraT, Pal, Oprl, Inp and AIDA-I.

14 For constructing the subject display systems, the first and second adapters can be homodimerization sequences or heterodimerization sequences. Preferred homodimerization sequences are two pairing cysteine residues capable of forming disulfide bond. Preferred heterodimerization sequences include those that are essentially incapable of forming homodimers under physiological buffer conditions and/or physiological body temperatures, such as those that are derived from heterodimeric receptors $GABA_B$ receptors 1 and 2. Other preferred adapters may adopt a coiled-coil secondary structure.

15 The present invention also provides a helper vector for displaying a polypeptide on the outer surface of a genetic package. The vector comprises: outer-surface sequences necessary for packaging the genetic package, wherein at least one of the surface presenting sequences is fused in-frame to an adapter, said adapter acting, when the polypeptide is produced in a suitable host cell, to cause the display of the polypeptide. In one aspect, the subject helper vector is a bacterial helper vector (see, e.g. Figure 26). In another aspect, the helper vector is a phage helper vector (also referred to herein as "UltraHelper phage vector"). Preferred phage helper vectors

include but are not limited to GM-UltraHelper phage vector shown in Figure 5A, CM-UltraHelper phage vector shown in Figure 13A, and GMCT-UltraHelper phage vector shown in Figure 19A.

16 The present invention further provides an expression vector for producing a polypeptide within or on the outer surface of a genetic package. The subject expression vector comprises: a coding sequence encoding the polypeptide fused in-frame to a first adapter, wherein the vector is devoid of outer-surface sequences encoding any functional outer-surface proteins of the genetic package, and display of the polypeptide on the outer surface of the genetic package is mediated via non-covalent pairwise interaction between the first adapter and a second adapter, wherein the second adapter is fused to an outer-surface sequence. In one aspect the expression vector is a phagemid. Illustrative phagemid of the subject phage display systems are pABMX14 shown in Figure 9A, pABMX15 shown in Figure 15A, and pAMBX22 shown in Figure 25A.

17 Also included in the present invention are kits comprising the adapter-directed display systems, and individual components of the systems including expression and helper vectors. Further provided in the invention are host cells comprising the subject vectors.

18 In a separate embodiment, the present invention provides a genetic package displaying on its external surface a fusion polypeptide. The fusion polypeptide comprises a polypeptide sequence to be displayed, fused in-frame with a first adapter, said first adapter acting, when the fusion polypeptide is produced in a suitable host cell, to cause the display of the fusion polypeptide via non-covalent pairwise interaction between the first adapter and a second adapter that is linked to an outer-surface protein. The genetic package can be viruses, cells, and spores.

19 In yet another embodiment, the present invention provides a selectable library comprising a plurality of genetic packages at least one being the genetic package as described above.

20 In still another embodiment, the present invention provides a method of displaying a polypeptide on the outer surface of a genetic package by causing the subject adapter-directed display system to be transcribed and translated in a suitable host cell. Selectable libraries produced by this method are also encompassed by the present invention.

21 The present invention further provides a method of detecting the presence of a specific interaction between a test agent and an exogenous polypeptide that is displayed on a genetic package. The method comprises the steps of (a) providing a genetic package displaying the exogenous polypeptide that is prepared according to the above-mentioned method; (b) contacting the genetic package with the test agent under conditions suitable to produce a stable polypeptide-agent complex; and (c) detecting the formation of the stable polypeptide-agent complex on the genetic package, thereby detecting the presence of a specific interaction. In one aspect, the exogenous polypeptide is selected from the group consisting of antigen-binding unit, cell surface receptor, receptor ligand, cytosolic protein, secreted protein, and nuclear protein. In a preferred aspect, the exogenous polypeptide is an antigen-binding unit. The test agent can be composed of protein, carbohydrate, lipid, and combinations thereof. Preferred test agent is an antigen or a ligand.

22 Finally, the present invention provides a method of obtaining a polypeptide with desired property. The method comprises: (a) providing a selectable library made by the invention method; and (b) screening the selectable library to obtain at least one genetic package displaying a polypeptide with the desired property. In one aspect, the desired property is binding specificity to an agent of interest. In another aspect, the step if

screening the selectable library further comprises isolating the genetic package that displays a polypeptide having the desired property. Such step may further involve obtaining a nucleotide sequence from the genetic package that encodes the polypeptide with the desired property. The polypeptide with the desired property may be one of the following types of proteins: antigen-binding unit, cell surface receptor, receptor ligand, cytosolic protein, secreted protein, and nuclear protein.

EXPLAINATION OF ABBREVIATIONS USED HEREIN

1. Nsc: Non-single chain
2. Sc: Sing-chain
3. Abu: Antigen-binding unit
4. Abus: Antigen-binding units
4. L chain: Light chain
5. H chain: Heavy chain
6. VL: Light chain variable region
7. VH: Heavy chain variable region

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic representation of the experimental design of the subject adapter-directed display systems. The depicted display systems not only permit expression of soluble exogenous polypeptides in a suitable host cell, but also allow display of the exogenous sequence on the outer surface of a genetic package. The system has two components: an expression vector and a helper vector. Introduction of the expression vector alone into a host cell such as *E. Coli* bacterium leads to expression

and secretion into the bacterial periplasm of the exogenous polypeptide which is fused in-frame with an adapter (designated "adapter 1," see center panel of Figure 1). Superinfection of the bacterial cells with a helper phage vector that carries a phage outer-surface sequence fused in-frame with a second adapter (designated "adapter 2," see left panel) permits display of the exogenous polypeptide on the phage particles via pairwise interaction between the first and second adapters. A diverse DNA sequences can be inserted into this expression vector to construct an expression library. Superinfection of the helper phages yields a diverse phage display libraries. Similarly, infection of the bacterial cells with the phagemid particles packaging a helper bacterial vector that carries a bacterial outer-surface sequence fused in-frame with a second adapter (designated "adapter 2," see right panel), permits display of the exogenous polypeptide on bacteria via pairwise interaction between the first and second adapters. A selectable bacterial display library can be constructed in a similar manner.

32 Figure 2 shows the results of a phage ELISA screen for kanamycin-resistant, KO7kpn helper phage-positive clones. 48 clones were screened for phage generation. C2, B3, B7, B9, A12 represent KO7kpn helper phage-positive clones. F1 and F2 represent two positive controls of parent M13KO7 phages.

33 Figure 3A is a schematic representation of the KO7kpn helper phage vector. Figure 3B depicts the nucleotide and amino acid sequence of the gene III leader sequence contained in the helper phage vector. A KpnI restriction site is introduced downstream of the leader sequence without altering coding region of gene III.

34 Figure 4 is a schematic representation of vector pABMC6. The vector contains KpnI site, a partial gene III leader sequence, GR2 coding sequence and a Myc- tag placed to the 5'of gene III sequence.

35 Figure 5A is a schematic representation of the GM-UltraHelper phage. Figure 5B depicts the nucleotide and amino acid sequences of the segment spanning the KpnI

and BgIII sites. Figure 5C depicts the trypsin cleavage sites (Tryp) in the GR2-Myc domain.

36 Figure 6 is a reproduction of an anti-Myc immunoblot of phage coat proteins. The results indicate the assembly of GM-UltraHelper phage particles. Lanes 1 and 7 represent molecular weight markers in kDa. Lanes 2-5 show four clones of GM-UltraHelper phages that express the GR2-Myc-pIII fusion. Lane 6 shows a negative control of M13KO7 helper phage that does not carry the GR2-Myc-pIII fusion sequence.

37 Figure 7 depicts the results of an ELISA assay using anti-Myc antibody to detect GR2-Myc- pIII fusion proteins that are assembled into the GM-UltraHelper phage particles. Phage clones 6, 18, and 20 shown in Figure 6 were tested. KO7 helper phage was included as a negative control.

38 Figure 8 depicts the results of an ELISA assay using anti-Myc antibody to demonstrate successful cleavage of GR2-Myc domain from UltraHelper phages using increasing amount of trypsin. M13KO7 helper phage served as a negative control.

39 Figure 9A is a schematic representation of vector pABMX14. Figure 9B shows the complete nucleotide sequence of pABMX14. The vector contains an ampicillin-resistance gene for antibiotic selection (AMP), a plasmid replication origin (ColE1 ori), the f1 phage replication origin (f1 ori), and the lac promoter/lac O1 driving the expression of downstream sequence plac-RBS-pelB-GR1-DH (HA and 6xHis tag). The NcoI/XbaI or NcoI/NotI or XbaI/NotI restriction sites can be used to insert exogenous sequence for display or production of soluble protein in a bacterial cell.

40 Figure 10 depicts the results of a phage binding assay, in which the phage particles were generated upon superinfection bacterial TG1 cells with either GM-UltraHelper phage or M13KO7 helper phage. The bacterial TG1 cells harbor the pABMX14-AM1 phagemid vector for expression of scFv-adapter 1 fusion. A dose-

dependent phage binding to the respective antigen was observed only upon the infection of GM-UltraHelper phage and not by the negative control M13KO7 phage. The results demonstrate the display of functional scFv on phage particles using pABMX14 phagemid and the GM-UltraHelper phage vector.

41 Figure 11, left panel, is a reproduction of an anti-Myc immunoblot of phage coat proteins. Lane 1 represents a negative control in which the phage particles were generated in TG1 cells harboring phagemid vector pABMX14-AM1 superinfected by M13KO7 helper phage. Lane 2 represents phage particles generated by phagemid pABMX14-AM1 vector and GM-UltraHelper phage vector. The exogenous sequence scFv was detected. Lane 3 represents a negative control in which GM-UltraHelper phage was employed alone. Anti-Myc antibody detects, only in lane 2, a band corresponding to the scFv-GR1-DH/GR2-Myc-pIII complex formed via the pairwise interaction of GR1 and GR2. Approximately twice as much scFv-GR1-DH/GR2-Myc-pIII complex was displayed when compared to the free GR2-Myc-pIII in line 2. This indicates that each phage particle on average carries more than one copy of the scFv-GR1 fusion. A band corresponding to a dimeric pIII-Myc-GR2 was detected in lane 3. The formation of pIII-Myc-GR2 dimer was due to the pair of cysteine residues introduced downstream of the $GABA_B$ receptor 2 adapter sequence. The adapter sequence of $GABA_B$ receptor 2 *per se* lacks the propensity to form homodimers under physiological body temperature and/or physiological buffer conditions (Kammerer *et al.* (1999) *Biochemistry* **38**: 13263-13269).

42 When the same blot was reprobed with the anti-HA antibodies (see right panel of Figure 11), a band corresponding to the scFv-GR1-DH/GR2-Myc-pIII complex was detected by the anti-HA antibodies in lane 2. This confirms the display of scFv-GR1-DH fusion upon the superinfection of GM-UltraHelper phages.

43 Figure 12 is a schematic representation of vector pABMC13. The vector contains a DNA fragment including a partial gIII leader sequence with KpnI site, Ala-Cys-Gly-Gly coding sequence and a Myc-tag placed to 5'of gene III.

44 Figure 13A is a schematic representation of the CM-UltraHelper phage vector. Figure 13B depicts the nucleotides and amino acids of the segment spanning the KpnI and BgIII sites. The vector contains a nucleotide sequence encoding Ala-Cys-Gly-Gly fused with a Myc-tag, which is placed to the 5'of gene III sequence. In addition, the vector comprises an amber stop codon flanked by the gene III leader sequence and the Cys-Myc coding region. The introduction of an amber codon permits phage production only in suppressor bacterial strains.

45 Figure 14 depicts the results of an ELISA assay using anti-Myc antibody to detect Cys-Myc- pIII fusions that are assembled into the CM-UltraHelper phage particles. Line 1 represents the negative control, M13KO7 helper phage. Lanes 2-6 represents five clones of CM-UltraHelper phages. Line 7 represents the positive control, GM-UltraHelper phage.

46 Figure 15A is a schematic representation of vector pABMX15. Figure 15B shows the complete nucleotide sequence of pABMX15. The vector contains an ampicillin-resistance gene for antibiotic selection (AMP), a plasmid replication origin (ColE1 ori), the f1 phage replication origin (f1 ori), and the lac promoter/lac O1 driving the expression of downstream sequence plac-RBS-pelB-HA-Cys. The NcoI/XbaI or NcoI/NotI or XbaI/NotI restriction sites can be used to insert exogenous sequence for display or production of soluble protein in a bacterial cell.

47 Figure 16 depicts the results of a phage binding assay, in which the phage particles were generated upon superinfection of bacterial TG1 cells with either CM-UltraHelper phage or M13KO7 (also denoted "KO7") helper phage. The bacterial TG1 cells harbors the pABMX15-AM1 phagemid vector for expression of scFv-HA-Cys

fusion. The results demonstrate the display of functional scFv on phage particles using pABMX15 phagemid and the CM-UltraHelper phage vector. There is no significant change in the level of scFv display in the range of 1 to 100 of multiplicity of infection (MIO) shown in line 3-6. Lines 1 and 2 represent two negative control of KO7 helper phages. As indicated by lane 2, no scFv-HA-Cys was detected when negative control M13KO7 helper phages were employed.

48 Figure 17, left panel, is a reproduction of an anti-Myc immunoblot of phage coat proteins. Lane 1 represents a negative control in which KO7 phage alone was employed. Lane 2 represents a negative control in which CM-UltraHelper phage vector alone was used. Lane 3 represents a negative control in which the scFv-HA-Cys fusion expressed by phagemid pABMX15 was not detected upon superinfection with KO7 phage. Lane 4 represents a phage clone in which the scFv-HA-Cys fusion expressed by phagemid pABMX15 was displayed successfully upon superinfection with CM-UltraHelper phage vector. Anti-Myc antibody detects, only in lane 4, a band corresponding to the scFv-HA-S-S-Myc-pIII complex formed via the pairwise interaction of the paring cysteines. This result indicates that display of scFv-HA-Cys occurs only when the phagemid is rescued by CM-UltraHelper phage and not by M13KO7 helper phage. pIII-Myc dimer was detected in line 2 and line 4 due to the disulfide bond established between two cysteine residues introduced at the 5' end of pIII sequence.

49 When the same blot was reprobed with the anti-HA antibodies (see right panel of Figure 17), a band corresponding to the scFv-HA-S-S-Myc-pIII complex was detected by the anti-HA antibodies in lane 4 and not in control lanes 1-3. This confirms the display of scFv-HA-S fusion upon the rescue of GM-UltraHelper phages.

50 Figure 18 is a schematic representation of vector pABMC12. In addition to the nucleotide sequence of vector pABMC6, vector pABMC12 contains a DNA fragment

including coding sequence for the C-terminal portion of gene III fused to GR2-Myc coding sequence, and a ribosome binding sequence (RBS)-OmpA leader sequence fused to a gene III sequence.

51 Figure 19A is a schematic representation of the GMCT-UltraHelper phage vector. Figure 19B depicts the nucleotides and amino acids of the segment spanning the KpnI and BgIII sites. The vector contains nucleotide sequence encoding the additional copy of engineered gene III fused to adapter GR2 and Myc-tag in KO7kpn phage vector, and ribosome binding sequence-OmpA leader sequence adjacent to the KO7 gene III sequence.

52 Figure 20 depicts the results of a phage binding assay, in which the phage particles were generated upon superinfection bacterial TG1 cells with either GMCT-UltraHelper phage or the control M13KO7 helper phage. The TG1 cells harbor the pABMX14-AM1 phagemid vector for expression of the scFv-GR1-DH fusion. The results demonstrate the display of functional scFv on phage particles using pABMX14 phagemid and the GMCT-UltraHelper phage vector (lanes 2-5). There is no significant change in the level of scFv display in the range of 1 to 100 of multiplicity of infection (MIO). Line 1 represents a negative control in which scFv was not displayed upon superinfection of KO7 helper phages.

53 Figure 21, left panel, is a reproduction of an anti-Myc immunoblot of phage coat proteins. Lane 1 represents a negative control in which KO7 phage was employed to rescue pABMX14 phagemid. Lane 2 represents a phage clone in which the scFv antibody expressed by phagemid pABMX14 was displayed successfully upon superinfection with GMCT-UltraHelper phages. Lane 3 represents another negative control in which GMCT-UltraHelper phage alone was used. Anti-Myc antibody detects, only in lane 2, a band corresponding to the scFv-GR1/GR2-Myc-CT (III) complex formed via the pairwise interaction of GR1 and GR2. This result indicates that display

of scFv-GR1 occurs only when the phagemid is rescued by GMCT -UltraHelper phage and not by KO7 helper phage.

54 When the same blot was reprobed with the anti-HA antibodies (see right panel of Figure 21), a band corresponding to the scFv-GR1/GR2-Myc-CT (III) complex was detected by the anti-HA antibodies in lane 2 and not in control lane 1 or 3. This confirms the display of scFv-HA-S fusion upon the rescue of GMCT-UltraHelper phages.

55 Figure 22A depicts a schematic representation of vector pABMD1 and pABMD2. Figure 22B depicts the nucleotide and amino acid sequences spanning the lac promoter/lac O1 and the SalI site.

56 Figure 23 depicts the C-terminal sequences of $GABA_B$ receptors 1 and 2. An exogenous cysteine residue is introduced by adding "ValGlyGlyCys" spacer at the C-termini of the sequence.

57 Figure 24 is a schematic representation depicting various antigen-binding units.

58 Figure 25A is a schematic representation of the bacterial expression vector pABMX22. Figure 25B depicts the complete nucleotide sequence of pABMX22. The vector contains an ampicillin-resistance gene for antibiotic selection (AMP), a plasmid replication origin (ColE1 ori), the f1 phage replication origin (f1 ori), and the lac promoter/lac O1 driving the expression of downstream sequence plac-RBS-p8L-GR1-HA. The MluI/XbaI or MluI/NotI or XbaI/NotI restriction sites can be used to insert exogenous sequence for display on a bacterial cell.

59 Figure 26A is a schematic representation of the bacterial helper vector pABMbd-1. Figure 26B depicts the complete nucleotide sequence of pABMbd-1. The vector contains a chloramphenicol-resistance gene for antibiotic selection (Cam), a plasmid replication origin (ColE1 ori), the f1 phage replication origin (f1 ori), and the lac

promoter/lac O1 driving the expression of downstream sequence plac-RBS-pelB-Lpp-OmpA-GR2.

MODE(S) FOR CARRYING OUT THE INVENTION

60 Throughout this disclosure, various publications, patents and published patent specifications are referenced by an identifying citation. The disclosures of these publications, patents and published patent specifications are hereby incorporated by reference into the present disclosure.

General Techniques:

61 The practice of the present invention will employ, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See, *e.g.,* PHAGE DISPLAY OF PEPTIDES AND PROTEINS (B.K. Kay et al., 1996); PHAGE DISPLAY, A LABORATORY MANUAL (C.F. Barbas III et al., 2001) Sambrook, Fritsch and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F. M. Ausubel, *et al.* eds., (1987)); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) ANTIBODIES, A LABORATORY MANUAL, and ANIMAL CELL CULTURE (R.I. Freshney, ed. (1987)).

62 As used in the specification and claims, the singular form "a", "an" and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a cell" includes a plurality of cells, including mixtures thereof.

Definitions:

63 The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear, cyclic, or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids. The terms also encompass amino acid polymers that have been modified, for example, via sulfation, glycosylation, lipidation, acetylation, phosphorylation, iodination, methylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, ubiquitination, or any other manipulation, such as conjugation with a labeling component. As used herein the term "amino acid" refers to either natural and/or unnatural or synthetic amino acids, including glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics.

64 A polypeptide or amino acid sequence "derived from" a designated protein refers to the origin of the polypeptide. Preferably, the polypeptide have an amino acid sequence that is essentially identical to that of a polypeptide encoded in the sequence, or a portion thereof wherein the portion consists of at least 10-20 amino acids, preferably at least 20-30 amino acids, more preferably at least 30-50 amino acids, or which is immunologically identifiable with a polypeptide encoded in the sequence. This terminology also includes a polypeptide expressed from a designated nucleic acid sequence.

65 A "chimeric" protein contains at least one fusion polypeptide comprising regions in a different position in the sequence than that occurs in nature. The regions may normally exist in separate proteins and are brought together in the fusion polypeptide; or they may normally exist in the same protein but are placed in a new arrangement in the fusion polypeptide. A chimeric protein may be created, for example, by chemical

synthesis, or by creating and translating a polynucleotide in which the peptide regions are encoded in the desired relationship.

66 A "multimeric protein" as used herein refers to a globular protein containing more than one separate polypeptide or protein chain associated with each other to form a single globular protein *in vitro* or *in vivo*. The multimeric protein may consist of more than one polypeptide of the same kind to form a "homomultimer." Alternatively, the multimeric protein may also be composed of more than one polypeptide of distinct sequences to form a "heteromultimer." Thus, a "heteromultimer" is a molecule comprising at least a first polypeptide and a second polypeptide, wherein the second polypeptide differs in amino acid sequence from the first polypeptide by at least one amino acid residue. The heteromultimer can comprise a "heterodimer" formed by the first and second polypeptide or can form higher order tertiary structures where more than two polypeptides are present. Exemplary structures for the heteromultimer include heterodimers (e.g. Fv and Fab fragments, diabodies, $GABA_B$ receptors 1 and 2 complexes), trimeric G-proteins, heterotetramers (e.g. $F(ab')_2$ fragments) and further oligomeric structures.

67 A "ligand" refers to a molecule capable of being bound by the ligand-binding domain of a receptor. The molecule may be chemically synthesized or may occur in nature.

68 An "agonist" is a molecule capable of stimulating the biological activity of a signaling molecule, e.g., a receptor.

69 An "antagonist" is a molecule capable of inhibiting the biological activity of a receptor.

70 By "pairwise interaction" is meant that the two adapters can interact with and bind to each other to form a stable complex. The stable complex must be sufficiently long-lasting to permit packaging the polypeptide onto the outer surface of a genetic

package. The complex or dimer must be able to withstand whatever conditions exist or are introduced between the moment of formation and the moment of detecting the displayed polypeptide, these conditions being a function of the assay or reaction which is being performed.

71 "Monovalent display" refers to expression of a single copy of the exogenous polypeptide per genetic package. In a monovalent phage display system, the collection of phage particles on average carries zero to one exogenous polypeptide per phage particle. By contrast, "multivalent display" refers to the expression of more than one copy of the exogenous polypeptide per genetic package. Thus, in a multivalent phage display system, the collection of phage particles on average carries more than one copy of the exogenous polypeptide.

72 The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen-binding site which specifically binds ("immunoreacts with") an antigen. Structurally, the simplest naturally occurring antibody (e.g., IgG) comprises four polypeptide chains, two heavy (H) chains and two light (L) chains inter-connected by disulfide bonds. The immunoglobulins represent a large family of molecules that include several types of molecules, such as IgD, IgG, IgA, IgM and IgE. The term "immunoglobulin molecule" includes, for example, hybrid antibodies, or altered antibodies, and fragments thereof. It has been shown that the antigen binding function of an antibody can be performed by fragments of a naturally-occurring antibody. These fragments are collectively termed "antigen-binding units" ("Abus"). Abus can be broadly divided into "single-chain" ("Sc") and "non-single-chain" ("Nsc") types based on their molecular structures.

73 Also encompassed within the terms "antibodies" and "Abus" are immunoglobulin molecules of a variety of species origins including invertebrates and

vertebrates. The term "human" as applies to an antibody or an Abu refers to an immunoglobulin molecule expressed by a human gene or fragment thereof. The term "humanized" as applies to a non-human (e.g. rodent or primate) antibodies are hybrid immunoglobulins, immunoglobulin chains or fragments thereof which contain minimal sequence derived from non-human immunoglobulin. For the most part, humanized antibodies are human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat, rabbit or primate having the desired specificity, affinity and capacity. In some instances, Fv framework region (FR) residues of the human immunoglobulin are replaced by corresponding non-human residues. Furthermore, the humanized antibody may comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. These modifications are made to further refine and optimize antibody performance and minimize immunogenicity when introduced into a human body. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin sequence. The humanized antibody may also comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin.

74 "Non-single-chain antigen-binding unit" ("Nsc Abus") are heteromultimers comprising a light-chain polypeptide and a heavy-chain polypeptide. Examples of the Nsc Abus include but are not limited to (1) a ccFv fragment (Figure 24) stabilized by the heterodimerization sequences disclosed herein; (2) any other monovalent and multivalent molecules comprising at least one ccFv fragment as described herein; (3) an Fab fragment consisting of the VL, VH, CL and CH1 domains; (4) an Fd fragment

consisting of the VH and CH1 domains; (5) an Fv fragment consisting of the VL and VH domains of a single arm of an antibody; (6) an F(ab')2 fragment, a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; (7) a diabody; and (8) any other Nsc Abus that are described in Little *et al.* (2000) Immunology Today.

75 As noted above, a Nsc Abus can be either "monovalent" or "multivalent." Whereas the former has one binding site per antigen-binding unit, the latter contains multiple binding sites capable of binding to more than one antigen of the same or different kind. Depending on the number of binding sites, a Nsc Abus may be bivalent (having two antigen-binding sites), trivalent (having three antigen-binding sites), tetravalent (having four antigen-binding sites), and so on.

76 Multivalent Nsc Abus can be further classified on the basis of their binding specificities. A "monospecific" Nsc Abu is a molecule capable of binding to one or more antigens of the same kind. A "multispecific" Nsc Abu is a molecule having binding specificities for at least two different antigens. While such molecules normally will only bind two distinct antigens (i.e. bispecific Abus), antibodies with additional specificities such as trispecific antibodies are encompassed by this expression when used herein. Examples of bispecific antigen binding units include those with one arm directed against a tumor cell antigen and the other arm directed against a cytotoxic trigger molecule such as anti-FcγRI/anti-CD15, anti-p185 ^{HER2} /FcγRIII (CD16), anti-CD3/anti-malignant B-cell (1D10), anti-CD3/anti-p185 ^{HER2}, anti-CD3/anti-p97, anti-CD3/anti-renal cell carcinoma, anti-CD3/anti-OVCAR-3, anti-CD3/L-D1 (anti-colon carcinoma), anti-CD3/anti-melanocyte stimulating hormone analog, anti-EGF receptor/anti-CD3, anti-CD3/anti-CAMA1, anti-CD3/anti-CD19, anti-CD3/MoV18, anti-neural cell adhesion molecule (NCAM)/anti-CD3, anti-folate binding protein (FBP)/anti-CD3, anti-pan carcinoma associated antigen (AMOC-31)/anti-CD3;

bispecific Abus with one arm which binds specifically to a tumor antigen and one arm which binds to a toxin such as anti-saporin/anti-Id-1, anti-CD22/anti-saporin, anti-CD7/anti-saporin, anti-CD38/anti-saporin, anti-CEA/anti-ricin A chain, anti-interferon-α (IFN-α)/anti-hybridoma idiotype, anti-CEA/anti-vinca alkaloid; BsAbs for converting enzyme activated prodrugs such as anti-CD30/anti-alkaline phosphatase (which catalyzes conversion of mitomycin phosphate prodrug to mitomycin alcohol); bispecific Abus which can be used as fibrinolytic agents such as anti-fibrin/anti-tissue plasminogen activator (tPA), anti-fibrin/anti-urokinase-type plasminogen activator (uPA); bispecific antigen-binding units for targeting immune complexes to cell surface receptors such as anti-low density lipoprotein (LDL)/anti-Fc receptor (e.g. Fcγ RI, FcγRII or FcγRIII); bispecific Abus for use in therapy of infectious diseases such as anti-CD3/anti-herpes simplex virus (HSV), anti-T-cell receptor:CD3 complex/anti-influenza, anti-FcγR/anti-HIV; bispecific Abus for tumor detection *in vitro* or *in vivo* such as anti-CEA/anti-EOTUBE, anti-CEA/anti-DPTA, anti-p185 ^{HER2} /anti-hapten; BsAbs as vaccine adjuvants; and bispecific Abus as diagnostic tools such as anti-rabbit IgG/anti-ferritin, anti-horse radish peroxidase (HRP)/anti-hormone, anti-somatostatin/anti-substance P, anti-HRP/anti-FITC, anti-CEA/anti-.beta.-galactosidase. Examples of trispecific antibodies include anti-CD3/anti-CD4/anti-CD37, anti-CD3/anti-CD5/anti-CD37 and anti-CD3/anti-CD8/anti-CD37.

77 Single-chain antigen-binding unit" ("Sc Abu") refers to a monomeric Abu. Although the two domains of the Fv fragment are coded for by separate genes, a synthetic linker can be made that enables them to be made as a single protein chain (i.e. single chain Fv ("scFv") as described in Bird *et al.* (1988) *Science* **242**:423-426 and Huston *et al.* (1988) *PNAS* **85**:5879-5883) by recombinant methods.

78 A "repertoire of antigen-binding units" refers to a plurality of antigen-binding units, at least two of which exhibit distinct binding specificities. A genetically diverse

repertoire of antigen-binding units refers to a plurality of antigen-binding units, the majority and if not all of the antigen-binding units exhibit unique binding specificities with respect to each other. Genetically diverse repertoire typically has a complexity of at least 10^6 to 10^{13}, preferably between 10^7 to 10^9, more preferably between 10^8 to 10^{10}, even more preferably between 10^8 to 10^{11} distinct antigen-binding units.

79 An antibody or Abu "specifically binds to" or "immunoreactive with" an antigen if it binds with greater affinity or avidity than it binds to other reference antigens including polypeptides or other substances.

80 An Abu is displayed "on the surface of a host cell" when the Abu is presented at the outer surface of a host cell. The displayed Abu may be directly attached to the outer surface of the host cell, or may be indirectly attached to the host cell via a host cell bound genetic package such as phage particle.

81 As used herein, "outer-surface sequences" refer to nucleotide sequences that encode "outer-surface proteins" of a genetic package. These proteins form a proteinaceous coat that encapsulates the genome of the genetic package. Typically, the outer-surface proteins direct the package to assemble the polypeptide to be displayed onto the outer surface of the genetic package, e.g. phage or bacteria.

82 The term "wildtype" as applied to a gene or a protein, refers to "naturally occurring," "native" gene or protein. These terms include full-length and processed polynucleotides and polypeptides that are naturally found in a cell or a genetic package. The term "wildtype outer-surface proteins" refers to those proteins forming the coat of naturally occurring genetic package, whether it is viruses, cells, or spores. In the case of filamentous bacteriophage, the wildtype proteins are gene III protein (pIII), gene VI protein (pVI), gene VII protein (pVII), gene VIII protein (pVIII), and gene IX protein (pIX).

83 "Antigen" as used herein means a substance that is recognized and bound specifically by an antibody. Antigens can include peptides, proteins, glycoproteins, polysaccharides and lipids; portions thereof and combinations thereof.

84 As used herein, the term "surface antigens" refers to the plasma membrane components of a cell. It encompasses integral and peripheral membrane proteins, glycoproteins, polysaccharides and lipids that constitute the plasma membrane. An "integral membrane protein" is a transmembrane protein that extends across the lipid bilayer of the plasma membrane of a cell. A typical integral membrane protein consists of at least one "membrane spanning segment" that generally comprises hydrophobic amino acid residues. Peripheral membrane proteins do not extend into the hydrophobic interior of the lipid bilayer and they are bound to the membrane surface by noncovalent interaction with other membrane proteins.

85 The terms "membrane", "cytosolic", "nuclear" and "secreted" as applied to cellular proteins specify the extracellular and/or subcellular location in which the cellular protein is mostly, predominantly, or preferentially localized.

86 "Cell surface receptors" represent a subset of membrane proteins, capable of binding to their respective ligands. Cell surface receptors are molecules anchored on or inserted into the cell plasma membrane. They constitute a large family of proteins, glycoproteins, polysaccharides and lipids, which serve not only as structural constituents of the plasma membrane, but also as regulatory elements, i.e., signaling molecules, governing a variety of biological functions.

87 A "heterodimeric receptor" encompasses cellular proteins composed of two proteinaceous subunits which exhibits binding affinity to a ligand. The two proteinaceous subunits are distinct molecules which differ in amino acid sequence by at least one amino acid residue. Non-limiting illustrative heterodimeric receptors are those that bind to growth factors (e.g. heregulin), neurotransmitters (e.g. γ-Aminobutyric

acid), and other organic or inorganic small molecules (e.g. mineralocorticoid, glucocorticoid). Preferred heterodimeric receptors are nuclear hormone receptors (Belshaw *et al.* (1996) *Proc. Natl. Acad. Sci. U. S. A* **93**(10):4604-4607), erbB3 and erbB2 receptor complex, and G-protein-coupled receptors including but not limited to opioid (Gomes *et al.* (2000) *J. Neuroscience* **20**(22): RC110); Jordan *et al.* (1999) *Nature* 399:697-700), muscarinic, dopamine, serotonin, adenosine/dopamine, and $GABA_B$ families of receptors.

88 "Domain" refers to a portion of a protein that is physically or functionally distinguished from other portions of the protein or peptide. Physically-defined domains include those amino acid sequences that are exceptionally hydrophobic or hydrophilic, such as those sequences that are membrane-associated or cytoplasm-associated. Domains may also be defined by internal homologies that arise, for example, from gene duplication. Functionally-defined domains have a distinct biological function(s). The ligand-binding domain of a receptor, for example, is that domain that binds ligand. An antigen-binding domain refers to the part of an antigen-binding unit or an antibody that binds to the antigen. Functionally-defined domains need not be encoded by contiguous amino acid sequences. Functionally-defined domains may contain one or more physically-defined domain. Receptors, for example, are generally divided into the extracellular ligand-binding domain, a transmembrane domain, and an intracellular effector domain. A "membrane anchorage domain" refers to the portion of a protein that mediates membrane association. Generally, the membrane anchorage domain is composed of hydrophobic amino acid residues. Alternatively, the membrane anchorage domain may contain modified amino acids, e.g. amino acids that are attached to a fatty acid chain, which in turn anchors the protein to a membrane.

89 A "host cell" includes an individual cell or cell culture which can be or has been a recipient for the subject vectors. Host cells include progeny of a single host cell. The

progeny may not necessarily be completely identical (in morphology or in genomic of total DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation. A host cell includes cells transfected *in vivo* with a vector of this invention.

90 A "cell line" or "cell culture" denotes bacterial, plant, insect or higher eukaryotic cells grown or maintained *in vitro*. The descendants of a cell may not be completely identical (either morphologically, genotypically, or phenotypically) to the parent cell.

91 A "defined medium" refers to a medium comprising nutritional and hormonal requirements necessary for the survival and/or growth of the cells in culture such that the components of the medium are known. Traditionally, the defined medium has been formulated by the addition of nutritional and growth factors necessary for growth and/or survival. Typically, the defined medium provides at least one component from one or more of the following categories: a) all essential amino acids, and usually the basic set of twenty amino acids plus cysteine; b) an energy source, usually in the form of a carbohydrate such as glucose; c) vitamins and/or other organic compounds required at low concentrations; d) free fatty acids; and e) trace elements, where trace elements are defined as inorganic compounds or naturally occurring elements that are typically required at very low concentrations, usually in the micromolar range. The defined medium may also optionally be supplemented with one or more components from any of the following categories: a) one or more mitogenic agents; b) salts and buffers as, for example, calcium, magnesium, and phosphate; c) nucleosides and bases such as, for example, adenosine and thymidine, hypoxanthine; and d) protein and tissue hydrolysates.

92 As used herein, the term "isolated" means separated from constituents, cellular and otherwise, in which the polynucleotide, peptide, polypeptide, protein, antibody, or fragments thereof, are normally associated with in nature. As is apparent to those of

skill in the art, a non-naturally occurring the polynucleotide, peptide, polypeptide, protein, antibody, or fragments thereof, does not require "isolation" to distinguish it from its naturally occurring counterpart. In addition, a "concentrated", "separated" or "diluted" polynucleotide, peptide, polypeptide, protein, antibody, or fragments thereof, is distinguishable from its naturally occurring counterpart in that the concentration or number of molecules per volume is greater than "concentrated" or less than "separated" than that of its naturally occurring counterpart.

93 Enrichment can be measured on an absolute basis, such as weight per volume of solution, or it can be measured in relation to a second, potentially interfering substance present in the source mixture. Increasing enrichments of the embodiments of this invention are increasingly more preferred. Thus, for example, a 2-fold enrichment is preferred, 10-fold enrichment is more preferred, 100-fold enrichment is more preferred, 1000-fold enrichment is even more preferred. A substance can also be provided in an isolated state by a process of artificial assembly, such as by chemical synthesis or recombinant expression.

94 "Linked" and "fused" or "fusion" are used interchangeably herein. These terms refer to the joining together of two more chemical elements or components, by whatever means including chemical conjugation or recombinant means. An "in-frame fusion" refers to the joining of two or more open reading frames (OFRs) to form a continuous longer OFR, in a manner that maintains the correct reading frame of the original OFRs. Thus, the resulting recombinant fusion protein is a single protein containing two or more segments that correspond to polypeptides encoded by the original OFRs (which segments are not normally so joined in nature.) Although the reading frame is thus made continuous throughout the fused segments, the segments may be physically or spatially separated by, for example, in-frame linker sequence (e.g. "flexon").

95 "Flexon" as used herein, refers to a flexible polypeptide linker (or a nucleic acid sequence encoding such a polypeptide) which typically comprises amino acids having small side chains (e.g. glycine, alanine, valine, leucine, isoleucine, and serine). Incorporating flexons between one or more sites of the subject fusions may promote functionality by allowing them to assume a conformations relatively independent of each other.

96 In the context of polypeptides, a "linear sequence" or a "sequence" is an order of amino acids in a polypeptide in an amino to carboxyl terminus direction in which residues that neighbor each other in the sequence are contiguous in the primary structure of the polypeptide. A "partial sequence" is a linear sequence of part of a polypeptide which is known to comprise additional residues in one or both directions.

97 "Heterologous" means derived from a genotypically distinct entity from the rest of the entity to which it is being compared. For example, a promoter removed from its native coding sequence and operatively linked to a coding sequence other than the native sequence is a heterologous promoter. The term "heterologous" as applied to a polynucleotide, a polypeptide, means that the polynucleotide or polypeptide is derived from a genotypically distinct entity from that of the rest of the entity to which it is being compared. For instance, a heterologous polynucleotide or antigen may be derived from a different species origin, different cell type, and the same type of cell of distinct individuals.

98 The terms "polynucleotides", "nucleic acids", "nucleotides" and "oligonucleotides" are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns,

messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component.

99 "Recombinant" as applied to a polynucleotide means that the polynucleotide is the product of various combinations of cloning, restriction and/or ligation steps, and other procedures that result in a construct that is distinct from a polynucleotide found in nature.

100 The terms "gene" or "gene fragment" are used interchangeably herein. They refer to a polynucleotide containing at least one open reading frame that is capable of encoding a particular protein after being transcribed and translated. A gene or gene fragment may be genomic or cDNA, as long as the polynucleotide contains at least one open reading frame, which may cover the entire coding region or a segment thereof.

101 "Operably linked" or "operatively linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. For instance, a promoter sequence is operably linked to a coding sequence if the promoter sequence promotes transcription of the coding sequence.

102 A "fusion gene" is a gene composed of at least two heterologous polynucleotides that are linked together.

103 A gene "database" denotes a set of stored data which represent a collection of sequences including nucleotide and peptide sequences, which in turn represent a collection of biological reference materials.

104 As used herein, "expression" refers to the process by which a polynucleotide is transcribed into mRNA and/or the process by which the transcribed mRNA (also referred to as "transcript") is subsequently being translated into peptides, polypeptides, or proteins. The transcripts and the encoded polypeptides are collectively referred to as gene product. If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell.

105 A "vector" is a nucleic acid molecule, preferably self-replicating, which transfers an inserted nucleic acid molecule into and/or between host cells. The term includes vectors that function primarily for insertion of DNA or RNA into a cell, replication of vectors that function primarily for the replication of DNA or RNA, and expression vectors that function for transcription and/or translation of the DNA or RNA. Also included are vectors that provide more than one of the above functions.

106 An "expression vector" is a polynucleotide which, when introduced into an appropriate host cell, can be transcribed and translated into a polypeptide(s). An "expression system" usually connotes a suitable host cell comprised of an expression vector that can function to yield a desired expression product.

107 A "replicon" refers to a polynucleotide comprising an origin of replication (generally referred to as an ori sequence) which allows for replication of the polynucleotide in an appropriate host cell. Examples of replicons include episomes (such as plasmids), as well as chromosomes (such as the nuclear or mitochondrial chromosomes).

108 "Signal transduction" is a process during which stimulatory or inhibitory signals are transmitted into and within a cell to elicit an intracellular response. A "modulator of a signal transduction pathway" refers to a compound which modulates the activity of one or more cellular proteins mapped to the same specific signal transduction pathway. A modulator may augment or suppress the activity of a signaling molecule.

Adapter-Directed Display System of the Present Invention

109 A central aspect of the present invention is the design of display systems that permit display on genetic packages of an exogenous polypeptide or a library of random or predetermined polypeptides, which is unlinked to any functional outer-surface sequences via peptide bonds. The subject systems avoid all drawbacks associated with expression of the outer-surface proteins by the expression vectors. The experimental design is particularly useful for presenting and/or selecting proteins with desired properties presented by genetic packages such as viruses, cells and spores.

110 The subject display systems comprise two components: (1) an expression vector that carries an exogenous gene of interest encoding a polypeptide to be displayed on the outer surface of a genetic package; and (2) a helper vector that facilitates the display of the polypeptide of particular interest. Distinguished from the previously reported display systems, the subject systems have the following unique features. First, the expression vector comprises a coding sequence encoding the exogenous polypeptide to be displayed fused in-frame with a first adapter. Second, the expression vector is devoid of outer-surface sequences that encode any functional outer-surface proteins of the genetic package. Third, the helper vector comprises all outer-surface sequences necessary for packaging the genetic package, at least one of the outer-surface sequences being fused in-frame to a second adapter sequence, and wherein the display of the exogenous polypeptide is mediated by pairwise interaction between the first and second adapters.

111 In one embodiment, the present invention provides a phage display system comprising a phagemid expression vector and a phage helper vector having the aforementioned characteristics. In another embodiment, the present invention provides a bacterial display system in which the bacterial expression vector and the bacterial

helper vector exhibit the claimed features. The experimental design of the phage and bacterial display systems can be extended to the construction of eukaryotic expression systems such as a mammalian cell display system.

Phage Display System of the Present Invention:

112 As noted above, previously reported phage display systems have a number of pronounced disadvantages. For instance, the commonly employed gene III and gene VIII systems bear several intrinsic drawbacks. Among them are (1) toxicity to the host cells as a result of expressing the exogenous polypeptide as a fusion with certain outer-surface proteins of the genetic package; (2) strict limitation on the size and orientation of the exogenous polypeptide to be displayed because certain regions of the outer-surface proteins are required for packaging the exogenous polypeptide onto the genetic package; and (3) instability of the fusion product due to recombination between the fusion and the wildtype outer-surface protein that is typically provided by a helper vector. The recently reported "cysteine-coupled" display system (WO 01/05950) avoids the expression of outer-surface protein fusions via peptide bond, but still fails to minimize the toxicity of these proteins to the host cells. Moreover, one particularly design, namely the two-vector system described in WO 01/05950, inevitably produces mispackaged vectors upon infection of the helper phages. The mispackaged vector contains the outer-surface sequences but not the exogenous gene. The subject phage display systems avoid these shortcomings and provide other related advantages.

113 A central aspect of the subject design is the separation of the exogenous polypeptide from the outer-surface proteins required for phage packaging. Thus, the phagemid vector (expression vector) carrying exogenous polypeptide does not contain any sequences that encode functional outer-surface proteins. The presentation of the exogenous polypeptide on the surface of the phage particle is mediated by the pairwise

interaction of two adapters. One of the adapters is fused in-frame with the exogenous polypeptide encoded by the phagemid vector, and the other is fused in-frame with at least one outer-surface proteins encoded by the helper phage vector. When the host cell carrying the phagemid vector is infected with the helper phage, the encoded exogenous polypeptide forms a complex with the outer-surface protein via pairwise interaction between the respective adapters. The complex is then packaged into surface sheath of the phage, leaving the exogenous polypeptide exposed on its outer surface.

General characteristics of the phagemid vectors of the present invention:

114 Several factors apply to the construction of the subject phage display systems. First, the phagemid vector does not contain sequences that encode any functional outer-surface proteins of the genetic package on which the polypeptide is to be displayed. By "functional" is meant that the encoded outer-surface proteins retain the ability to facilitate or direct the genetic package to assemble the polypeptide of interest onto its outer surface. The precise outer-surface sequences to be excluded from the expression vectors will depend on the choice of phage packages.

115 As used herein, the term "phage" encompasses viruses consisting of a protein coat encapsulated therein a viral genome required for viral replication. The viral genome may be composed of DNA or RNA, single or double stranded, linear or circular. The phages may infect a wide range of host cells, including but not limited to prokaryotes such as bacterial cells. The genomes of many phages, filamentous or non-filamentous, have been sequenced. Representative filamentous phages include M13, f1, fd, Ifl, Ike, Xf, Pf1, and Pf3. Within the class of filamentous phages, M13 is the most well-characterized species. Its 3-dimentional structure is known, and the functions of its coat proteins are well understood. Specifically, the M13 genome encodes five coat proteins, namely pIII, VIII, VI, VII and IX. For constructing an M13-based expression

vector of the subject phage display system, all of the coat-encoding sequences must be deleted, or altered so that the encoded protein products are incapable of effecting the presentation of the exogenous polypeptide onto the outer surface of a phage particle. Suitable modifications to a functional outer-surface protein may result in: (1) loss of functional signal peptide that directs the intracellular translocation of the outer-surface protein into the periplasm of the bacterial cells, where the signal peptide is then cleaved off; (2) loss of function of the coat protein domain that anchors the mature polypeptide into the bacterial cell membrane and/or phage coat; (3) loss of function of the coat protein domain that specifically binds to the phage receptor, the F-pilus of the host bacterium; and/or (4) introduction of internal stop codons to prevent expression of any functional coat proteins. These and other domains within several coat proteins, such as pIII, have been delineated (see, e.g. U.S. Patent No. 5,969,108). The outer-surface proteins of other closely related members such as f1 and fd filamentous phages are also well known in the art (see, e.g. Kay et al. (1996) Phage Display of Peptides and Protiens: A Laboratory Manual. Academic Press., Inc. San Diego). Preferably, the only phage sequence presented in an M13-based expression vector contains f1 origin required for phagemid replication and package. A stepwise illustration on constructing an M13-based expression vector is detailed in Examples 1-4. Thus, one of ordinary skill in the art can readily construct an expression vector with the claimed features without undue experimentation.

116 Similar constructions can be made with other filamentous phage. Pf3 is another well-known filamentous phage that infects *Pseudomonas aerugenosa* cells that harbor an IncP-1 plasmid. The entire genome of Pf3 has been sequenced and the genetic signals involved in replication and assembly have been characterized (Luiten et al. (1985) *J. Virology* **56** (1): 268-276). The major coat protein of Pf3 is unusual in having no signal peptide to direct its secretion. The sequence has charged residues ASP_7,

ARG_{37}, LYS_{40}, and PHE_{44}-COO^- which is consistent with the amino terminus being exposed. The viral strand replication origin of 139 bp DNA for Pf3 phage has also been identified (Luiten et al. (1991) *J. Bacteriol* **173**(13): 4007-4012). To construct a Pf3-based expression vector, the Pf3 coat-encoding sequence must be deleted or altered so that no functional major coat protein is encoded. A preferred expression vector only contains the Pf3 phage replication origin for its replication and packaging.

117 The same approach applies to construction of phagemid vectors derived from non-filamentous phages. Non-limiting representative members of this class of phages are bacteriophage ϕX174, λ, T4 and T7. The bacteriophage ϕX174 is a very small icosahedral virus which has been thoroughly studied by genetics, biochemistry, and electron microscopy. Three gene products of ϕX174 are present on the outside of the mature virion: F (capsid), G (major spike protein, 60 copies per virion), and H (minor spike protein, 12 copies per virion). The G protein comprises 175 amino acids, while H comprises 328 amino acids. The F protein interacts with the single-stranded DNA of the virus. The proteins F, G, and H are translated from a single mRNA in the viral infected cells. Thus, an exemplary expression vector based on this class of non-filamentous phage lacks all coding sequences of the F, G and H proteins. Other alternative expression vectors comprise altered F-, G-, or H-encoding sequences that do not yield functional F, G and H proteins.

General characteristics of the helper phage vectors of the present invention:

118 The second component of the subject phage display systems is a helper vector that functions to complement the expression vectors devoid of any functional outer-surface sequences. Unlike previously described helper vectors (U.S. Patent No. 5,969,108) that either lack one of the necessary coat-protein encoding sequences, or contain a sequence encoding a defective coat protein of the genetic package, the subject

helper vectors provide all of the outer-surface sequences required for packaging the genetic package. The precise outer-surface sequences employed again depend on the choice of the phage packages.

119 As mentioned above, a wealth of structural and biochemical information on a variety of phages is available in the art. The gene sequences encoding structural proteins and enzymes required for replicating and packaging numerous types of genetic package have been identified, and widely used to construct prior display systems (U.S. Patent Nos. 6248516, 5969108, 5885793, 5837500, 5571698, 5223409, 5514548, WO9005144, EP0368684, WO09201047, WO09311236, and WO09708320). These sequences are generally applicable for constructing the subject helper vectors exhibiting additional unique features.

120 Specifically, the subject helper phage vector generally comprises all outer-surface sequences responsible for encapsulating both the helper phage and the phagemid vector. In one aspect, the helper phage vector comprises outer-surface sequences encoding all coat proteins of one of the following filamentous phages M13, f1, fd, Ifl, Ike, Xf, Pf1, and Pf3. Preferred coat-encoding sequences of an M13-based helper phage vector are gIII, gVIII, gVI, gVII, gIX or their functional equivalents. In another aspect, the helper phage vector contains coat encoding sequences of a non-filamentous phage selected from the group consisting of bacteriophage ϕX174, λ, T4 and T7. In addition to these structural proteins, the helper phage vector typically encodes other phage-derived enzymes that act in *trans* on the phage origins of replication carried on both the phagemid vector and helper phage to "help" replicate and package the phagemid vector. A preferred M13 helper vector of the present invention is replication-defective, so as to ensure preferential packaging of the phagemid vectors. Preferably, more than 90% of the packaged vectors are phagemid vectors; even more preferably more than 99% of the packaged vectors are phagemid vectors. The preferred M13 helper phage vector further

comprises sequences encoding proteins I, II, IV, V, X or functional equivalents. As used herein, the functional equivalents of outer-surface proteins include those coding for modified outer-surface proteins that retain the functionality of the wildtype outer-surface proteins. Functionally equivalent outer-surface proteins include those that enhance, decrease or not significantly affect properties of the corresponding wildtype proteins. These equivalents may be polypeptides having conservative amino acid substitutions, analogs including fusions and mutants. A preferred M13 helper vector is interference-resistant. Exemplary interference-resistant helper vectors are M13K07 (Amersham Pharmacia Biotech) and in its derivatives such as VCSM13 (Stratagene). These two interference-resistant helper vectors provide the phage sequences necessary for packaging a phage particle.

121 The helper phage vector of the present invention may contain one or more copies of a given outer-surface sequence, so long as all of the outer-surface sequences necessary for phage packaging are present. The use of one copy of each necessary outer-surface sequence typically yields a "multivalent" phage display system. By contrast, the incorporation of more than one copy of a given outer-surface sequence or its functional equivalent potentially yields a "monovalent" phage display system. Monovalent display allows the discrimination of displayed polypeptides that bind targets with moderate versus high affinity. It also aids in selection of polypeptides such as antibodies, on the basis of affinity by avoiding the "avidity" effect where a phage expressing more than one copy of a low affinity antibody would have the same apparent affinity as a phage expressing one copy of a higher affinity antibody. Multivalent display, however, provides alternative advantages. It is particularly useful in the initial stages of selection of binding polypeptides. At the early stages of screening, it is often preferably to accumulate a broad spectrum of polypeptides as potential leads than to identify a single high-affinity candidate. The polypeptides obtained through the initial

screen can then be ordered in terms of their affinities using monovalent display or other methods. The present invention provides an exemplary helper phage vector (Figure 5A) of a multivalent display system. Upon infecting bacterial cells with the resulting helper phages, the packaged phagemid particles exhibit about two fold more exogenous polypeptide/pIII complex than the free pIII, indicating that the packaged phages have a valency of more than one copy (see Example 2, Figure 11). Also provided in the present invention is an M13 helper phage vector carrying a copy of the KO7 gene III outer-surface sequence and a copy the C-terminal portion of gene III (Figure 19A and Example 4). The latter sequence encodes a functional equivalent capable of competing with the wildtype pIII for packaging. The resulting helper phages are expected to yield a monovalent display system.

122 A particularly preferred helper vector supports both monovalent and multivalent display. Such a helper vector can be constructed by incorporating a suppressible translational stop codon between the first and second copy of an outer-surface sequence, e.g. gene III of M13 phage. The suppressible codon allows the translation of nucleotide sequences downstream of the codon (e.g. gene III) under suppressive condition, but under non-suppressive conditions translation ends at the codon. When the helper phage is grown in suppressive condition, for example in suppressor bacterial strains, the second copy of the outer-surface protein is expressed which competes with the first copy for packaging; a monovalent phage display occurs. However, when the same helper phage is grown in a non-suppressive bacterial strain, the second copy of the outer-surface sequence will not be expressed, and thus yield a multivalent display system. As such, the suppressible codon functions as a convenient "switch" that controls either form of display when subjected to two different conditions. Examples of suppressible translational stop codons are the amber, ochre and opal codons.

General characteristics of the adapters of the present invention:

123 A further consideration in constructing the phage display system is to select a pair of adapter sequences that encode two adapters capable of pairwise interaction. Whereas one of the adapter sequences is inserted in-frame with the exogenous sequence carried by the phagemid vector, the other is fused in-frame with at least one of the outer-surface protein of the helper phage vector. By "pairwise interaction" is meant that the two adapters can interact with and bind to each other to form a stable complex. The stable complex must be sufficiently long-lasting to permit packaging the polypeptide onto the outer surface of the genetic package. The complex or dimer must be able to withstand whatever conditions exist or are introduced between the moment of formation and the moment of detecting the displayed polypeptide, these conditions being a function of the assay or reaction which is being performed. For phages (e.g. M13) that are assembled periplasmically, the complex or dimer must be sufficiently stable when residing in the bacterial periplasm, where it is packaged along with the phage genome. The stable complex or dimer may be irreversible or reversible as long as it meets the other requirements of this definition. Thus, a transient complex or dimer may form in a reaction mixture, but it does not constitute a stable complex if it dissociates spontaneously and yields no detectable polypeptide displayed on the outer surface of a genetic package.

124 The pairwise interaction between the first and second adapters may be covalent or non-covalent interactions. Non-covalent interactions encompass every exiting stable linkage that do not result in the formation of a covalent bond. Non-limiting examples of noncovalent interactions include electrostatic bonds, hydrogen bonding, Van der Waal's forces, steric interdigitation of amphiphilic peptides. By contrast, covalent interactions result in the formation of covalent bonds, including but not limited to disulfide bond between two cysteine residues, C-C bond between two carbon-containing molecules, C-

O or C-H between a carbon and oxygen- or hydrogen-containing molecules respectively, and O-P bond between an oxygen- and phosphate-containing molecule.

125 Adapter sequences applicable for constructing the expression and helper vectors of the subject display system can be derived from a variety of sources. Generally, any protein sequences involved in the formation of stable multimers are candidate adapter sequences. As such, these sequences may be derived from any homomultimeric or heteromultimeric protein complexes. Representative homomultimeric proteins are homodimeric receptors (e.g. platelet-derived growth factor homodimer BB (PDGF), homodimeric transcription factors (e.g. Max homodimer, NF-kappaB p65 (RelA) homodimer), and growth factors (e.g. neurotrophin homodimers). Non-limiting examples of heteromultimeric proteins are complexes of protein kinases and SH2-domain-containing proteins (Cantley *et al.* (1993) *Cell* **72**: 767-778; Cantley *et al.* (1995) *J. Biol. Chem.* **270**(44): 26029-26032), heterodimeric transcription factors, and heterodimeric receptors.

126 Preferred heterodimeric transcription factors are a-Pal/ Max complexes and Hox/Pbx complexes. Hox represents a large family of transcription factors involved in patterning the anterior-posterior axis during embryogenesis. Hox proteins bind DNA with a conserved three alpha helix homeodomain. In order to bind to specific DNA sequences, Hox proteins require the presence of hetero-partners such as the Pbx homeodomain. Wolberger *et al.* solved the 2.35Å crystal structure of a HoxB1-Pbx1-DNA ternary complex in order to understand how Hox-Pbx complex formation occurs and how this complex binds to DNA. The structure shows that the homeodomain of each protein binds to adjacent recognition sequences on opposite sides of the DNA. Heterodimerization occurs through contacts formed between a six amino acid hexapeptide N-terminal to the homeodomain of HoxB1 and a pocket in Pbx1 formed between helix 3 and helices 1 and 2. A C-terminal extension of the Pbx1 homeodomain

forms an alpha helix that packs against helix 1 to form a larger four helix homeodomain (Wolberger *et al.* (1999) *Cell* **96**: 587-597; Wolberger *et al. J Mol Biol.* **291**: 521-530).

127 A vast number of heterodimeric receptors have also been identified. They include but are not limited to those that bind to growth factors (e.g. heregulin), neurotransmitters (e.g. γ-Aminobutyric acid), and other organic or inorganic small molecules (e.g. mineralocorticoid, glucocorticoid). Preferred heterodimeric receptors are nuclear hormone receptors (Belshaw *et al.* (1996) *Proc. Natl. Acad. Sci. U. S. A* **93**(10):4604-4607), erbB3 and erbB2 receptor complex, and G-protein-coupled receptors including but not limited to opioid (Gomes *et al.* (2000) *J. Neuroscience* **20**(22): RC110); Jordan *et al.* (1999) *Nature* 399:697-700), muscarinic, dopamine, serotonin, adenosine/dopamine, and $GABA_B$ families of receptors. For majority of the known heterodimeric receptors, their C-terminal sequences are found to mediate heterodimer formation.

128 Sequence of antibody chains that are involved in dimerizing the L and H chains can also be used as adapters for constructing the subject display systems. These sequences include but are not limited to constant region sequences of an L or H chain. Additionally, adapter sequences can be derived from antigen-binding site sequences and its binding antigen. In such case, one adapter of the pair contains antigen-binding site amino acid residues that is recognized (i.e. being able to stably associate with) by the other adapter containing the corresponding antigen residues.

129 Based on the wealth of genetic and biochemical data on vast families of genes, one of ordinary skill will be able to select and obtain suitable adapter sequences for constructing the subject display system without undue experimentation.

130 Where desired, sequences from novel hetermultimeric proteins can be employed as adapters. In such situation, the identification of candidate sequences involved in formation of heteromultimers can be determined by any genetic or biochemical assays

without undue experimentation. Additionally, computer modeling and searching technologies further facilitates detection of heteromultimeric sequences based on sequence homologies of common domains appeared in related and unrelated genes. Non-limiting examples of programs that allow homology searches are Blast (http://www.ncbi.nlm.nih.gov/BLAST/), Fasta (Genetics Computing Group package, Madison, Wisconsin), DNA Star, Clustlaw, TOFFEE, COBLATH, Genthreader, and MegAlign. Any sequence databases that contains DNA sequences corresponding to a target receptor or a segment thereof can be used for sequence analysis. Commonly employed databases include but are not limited to GenBank, EMBL, DDBJ, PDB, SWISS-PROT, EST, STS, GSS, and HTGS.

131 The subject adapters that are derived from heterodimerization sequences can be further characterized based on their physical properties. Preferred heterodimerization sequences exhibit pairwise affinity resulting in predominant formation of heterodimers to a substantial exclusion of homodimers. Preferably, the predominant formation yields a heteromultimeric pool that contains at least 60% heterodimers, more preferably at least 80% heterodimers, more preferably between 85-90% heterodimers, and more preferably between 90-95% heterodimers, and even more preferably between 96-99% heterodimers that are allowed to form under physiological buffer conditions and/or physiological body temperatures. In certain embodiments of the present invention, at least one of the heterodimerization sequences of the adapter pair is essentially incapable of forming a homodimer in a physiological buffer and/or at physiological body temperature. By "essentially incapable" is meant that the selected heterodimerization sequences when tested alone do not yield detectable amounts of homodimers in an *in vitro* sedimentation experiment as detailed in Kammerer *et al.* (1999) *Biochemistry* **38**: 13263-13269), or in the *in vivo* two-hybrid yeast analysis (see e.g. White *et al. Nature* (1998) **396**: 679-682). In addition, individual heterodimerization sequences can be expressed in a host cell and

the absence of homodimers in the host cell can be demonstrated by a variety of protein analyses including but not limited to SDS-PAGE, Western blot, and immunoprecipitation. The *in vitro* assays must be conducted under a physiological buffer conditions, and/or preferably at physiological body temperatures. Generally, a physiological buffer contains a physiological concentration of salt and at adjusted to a neutral pH ranging from about 6.5 to about 7.8, and preferably from about 7.0 to about 7.5. A variety of physiological buffers is listed in Sambrook *et al.* (1989) *supra* and hence is not detailed herein. Preferred physiological conditions are described in Kammerer *et al.*, *supra*.

132 An illustrative adapter pair exhibiting the above-mentioned physical properties is $GABA_B$-R1/$GABA_B$-R2 receptors. These two receptors are essentially incapable of forming homodimers under physiological conditions (e.g. *in vivo*) and at physiological body temperatures. Research by Kuner *et al.* and White *et al.* (*Science* (1999) **283**: 74-77); *Nature* (1998) **396**: 679-682)) has demonstrated the heterodimerization specificity of $GABA_B$-R1 and $GABA_B$-R2 *in vivo*. In fact, White *et al.* were able to clone $GABA_B$-R2 from yeast cells based on the exclusive specificity of this heterodimeric receptor pair. *In vitro* studies by Kammerer *et al. supra* has shown that neither $GABA_B$-R1 nor $GABA_B$-R2 C-terminal sequence is capable of forming homodimers in physiological buffer conditions when assayed at physiological body temperatures. Specifically, Kammerer et al. have demonstrated by sedimentation experiments that the heterodimerization sequences of $GABA_B$ receptor 1 and 2, when tested alone, sediment at the molecular mass of the monomer under physiological conditions and at physiological body temperatures (e.g. at 37^0C). When mixed in equimolar amounts, $GABA_B$ receptor 1 and 2 heterodimerization sequences sediment at the molecular mass corresponding to the heterodimer of the two sequences (see Table 1 of Kammerer et al.).

However, when the $GABA_B$-R1 and $GABA_B$-R2 C-terminal sequences are linked to a cysteine residue, homodimers may occur via formation of disulfide bond.

133 Adapters can be further characterized based on their secondary structures. Preferred adapters consist of amphiphilic peptides that adopt a coiled-coil helical structure. The helical coiled coil is one of the principal subunit oligomerization sequences in proteins. Primary sequence analysis reveals that approximately 2-3% of all protein residues form coiled coils (Wolf *et al.* (1997) *Protein Sci.* **6**:1179-1189). Well-characterized coiled-coil-containing proteins include members of the cytoskeletal family (e.g. α-keratin, vimentin), cytoskeletal motor family (e.g. myosine, kinesins, and dyneins), viral membrane proteins (e.g. membrane proteins of Ebola or HIV), DNA binding proteins, and cell surface receptors (e.g. $GABA_B$ receptors 1 and 2). Coiled-coil adapters of the present invention can be broadly classified into two groups, namely the left-handed and right-handed coiled coils. The left-handed coiled coils are characterized by a heptad repeat denoted "abcdefg" with the occurrence of apolar residues preferentially located at the first (a) and fourth (d) position. The residues at these two positions typically constitute a zig-zag pattern of "knobs and holes" that interlock with those of the other stand to form a tight-fitting hydrophobic core. In contrast, the second (b), third (c) and sixth (f) positions that cover the periphery of the coiled coil are preferably charged residues. Examples of charged amino acids include basic residues such as lysine, arginine, histidine, and acidic residues such as aspartate, glutamate, asparagine, and glutamine. Uncharged or apolar amino acids suitable for designing a heterodimeric coiled coil include but are not limited to glycine, alanine, valine, leucine, isoleucine, serine and threonine. While the uncharged residues typically form the hydrophobic core, inter-helical and intra-helical salt-bridge including charged residues even at core positions may be employed to stabilize the overall helical coiled-coiled structure (Burkhard *et al.* (2000) *J. Biol. Chem.* **275**:11672-11677). Whereas varying

lengths of coiled coil may be employed, the subject coiled coil adapters preferably contain two to ten heptad repeats. More preferably, the adapters contain three to eight heptad repeats, even more preferably contain four to five heptad repeats.

134 In designing optimal coiled-coil adapters, a variety of existing computer software programs that predict the secondary structure of a peptide can be used. An illustrative computer analysis uses the COILS algorithm which compares an amino acid sequence with sequences in the database of known two-stranded coiled coils, and predicts the high probability coiled-coil stretches (Kammerer *et al.*(1999) *Biochemistry* **38**:13263-13269).

135 While a diverse variety of coiled coils involved in multimer formation can be employed as the adapters in the subject display system. Preferred coiled coils are derived from heterodimeric receptors. Accordingly, the present invention encompasses coiled-coil adapters derived from $GABA_B$ receptors 1 and 2. In one aspect, the subject coiled coils adapters comprise the C-terminal sequences of $GABA_B$ receptor 1 and $GABA_B$ receptor 2. In another aspect, the subject adapters are composed of two distinct polypeptides of at least 30 amino acid residues, one of which is essentially identical to a linear sequence of comparable length depicted in Figure 23 (GR1), and the other is essentially identical to a linear peptide sequence of comparable length depicted in Figure 23 (GR2).

136 Another class of preferred coiled coil adapters are leucine zippers. The leucine zipper have been defined in the art as a stretch of about 35 amino acids containing 4-5 leucine residues separated from each other by six amino acids (Maniatis and Abel, (1989) *Nature* **341**:24). The leucine zipper has been found to occur in a variety of eukaryotic DNA-binding proteins, such as GCN4, C/EBP, c-fos gene product (Fos), c-jun gene product (Jun), and c-Myc gene product. In these proteins, the leucine zipper creates a dimerization interface wherein proteins containing leucine zippers may form

stable homodimers and/or heterodimers. Molecular analysis of the protein products encoded by two proto-oncogenes, c-fos and c-jun, has revealed such a case of preferential heterodimer formation (Gentz et al., (1989) *Science* **243**:1695; Nakabeppu et al., (1988) *Cell* **55**:907; Cohen et al., (1989) *Genes Dev.* **3**:173). Synthetic peptides comprising the leucine zipper regions of Fos and Jun have also been shown to mediate heterodimer formation, and, where the amino-termini of the synthetic peptides each include a cysteine residue to permit intermolecular disulfide bonding, heterodimer formation occurs to the substantial exclusion of homodimerization.

137 The leucine-zipper adapters of the present invention have the general structural formula known as the heptad repeat (Leucine- $X_1 - X_2 - X_3 - X_4 - X_5 - X_6)_n$, where X may be any of the conventional 20 amino acids, but are most likely to be amino acids with alpha-helix forming potential, for example, alanine, valine, aspartic acid, glutamic acid, and lysine, and n may be 2 or greater, although typically n is 3 to 10, preferably 4 to 8, more preferably 4 to 5. Preferred sequences are the Fos or Jun leucine zippers.

138 As used herein, a linear sequence of peptide is "essentially identical" to another linear sequence, if both sequences exhibit substantial amino acid or nucleotide sequence homology. Generally, essentially identical sequences are at least about 60% identical with each other, after alignment of the homologous regions. Preferably, the sequences are at least about 70% identical; more preferably, they are at least about 80% identical; more preferably, they are at least about 90% identical; more preferably, the sequences are at least about 95% identical; still more preferably, the sequences are 100% identical.

139 In determining whether polypeptide sequences are essentially identical, a sequence that preserves the functionality of the polypeptide with which it is being compared is particularly preferred. Functionality may be established by different criteria, such as ability to form a stable complex with a pairing adapter, and ability to facilitate display of polypeptides fused in-frame with the adapter.

140 The subject adapters include modified leucine zippers and $GABA_B$ heterodimerization sequences which are functionally equivalent to the polypeptide sequences exemplified herein. Modified polypeptides providing improved stability to the paired adapters and/or display efficiency are preferred. Examples of modified polypeptides include those with conservative substitutions of amino acid residues, and one or more deletions or additions of amino acids which do not significantly deleteriously alter the heterodimerization specificity. Substitutions can range from changing or modifying one or more amino acid residues to complete redesign of a region as long as the pairwise interaction is maintained. Amino acid substitutions, if present, are preferably conservative substitutions that do not deleteriously affect folding or functional properties of the peptide. Groups of functionally related amino acids within which conservative substitutions can be made are glycine/alanine; valine/isoleucine/leucine; asparagine/glutamine; aspartic acid/glutamic acid; serine/threonine/methionine; lysine/arginine; and phenylalanine/tryosine/tryptophan. Polypeptides of this invention can be in glycosylated or unglycosylated form, can be modified post-translationally (e.g., acetylation, and phosphorylation) or can be modified synthetically (e.g., the attachment of a labeling group).

141 The adapter sequences of the present invention can be obtained using conventional recombinant cloning methods and/or by chemical synthesis. Using well-established restriction and ligation techniques, the appropriate adapter sequences can be excised from various DNA sources and integrated in-frame with the exogenous gene sequences and the outer-surface sequences to generate the expression and helper vectors, respectively.

142 Preferably, the first adapter sequence is inserted into the expression vector in such a way to minimize structural interference, if any, on the resulting exogenous fusion polypeptide. Whereas the first adapter can be fused to the 5' or 3' of the exogenous

gene sequence, Figure 9A depicts a preferred phagemid vector in which the adapter sequence (i.e. hererodimerization sequence derived from $GABA_B$ receptor 1) is fused in-frame to the 3' end of the exogenous gene sequence.

143 Similarly, the second adapter sequence is inserted into the helper vector in a position where the integrity of the expressed phage coat is not undermined. The adapter sequence can be fused to the 5' or 3' end of an outer-surface sequence without disrupting the coding region. Figures 5 and 19 depict two preferred helper phage vectors in which the adapter sequence (i.e. heterodimerizeration sequence derived from $GABA_B$ receptor 2) is placed in-frame to the 5' end of the outer-surface sequence, gene III or a functional portion thereof.

Bacterial Display System of the Present Invention:

144 The present invention also provides a bacterial display system comprising the following two components: (1) a bacterial expression vector that carries an exogenous gene of interest encoding an exogenous polypeptide to be displayed on the outer surface of a bacterial cell or bacterial spore; and (2) a helper vector that facilitates the display of the polypeptide of particular interest. Unlike the previously described bacterial systems in which the exogenous polypeptide is expressed as a fusion with a bacterial outer-surface protein, the subject bacterial systems have the following unique features. First, the bacterial expression vector comprises a coding sequence encoding the exogenous polypeptide to be displayed fused in-frame with a first adapter. Second, the bacterial expression vector is devoid of outer-surface sequences that encode functional outer-surface proteins of a bacterial or bacterial spore. Third, the helper vector comprises all outer-surface sequences necessary for packaging the genetic package, at least one of the outer-surface sequences being fused in-frame to a second adapter sequence, and wherein

the display of the exogenous polypeptide is mediated by pairwise interaction between the first and second adapters.

145 The general principle and experimental design outlined above for constructing the subject phage display system are equally applicable for generating the subject bacterial display system. Whereas the bacterial expression vector lacks sequences encoding any functional outer-surface proteins, the helper bacterial vector contains an outer-surface sequence necessary for compensating the deficiency.

146 The helper bacterial vector typically comprises an outer-surface sequence encoding an outer-surface protein having the following two domains: (1) a signal peptide that directs the protein to be secreted through the lipid bilayer to the periplasm; and (2) a membrane translocating domain capable of locating the outer-surface protein onto the outer surface of a bacterial cell. The expressed outer-surface protein is first transported to the periplasm where the leader peptide is cleaved off. When the outer-surface protein is expressed as a fusion with an adapter, the adapter facilitates the translocation of the exogenous polypeptide that is also present in the periplasm onto the bacterial outer surface upon binding to the paring adapter contained in the exogenous polypeptide.

147 Prior research has revealed a vast number of bacterial surface-protein encoding sequences that can be used for constructing the helper bacterial expression vector. Non-limiting examples of bacterial surface proteins are LamB (Bremer *et al. Proc. Natl. Acad. Sci U.S.A.* (1984) **81**:3830-34; *Gene* (1987) **52:**165-73); OmpA (*Prog Biophys Molec Biol* (1987) **49**:89-115); OmpC (Misra et al. (1988) *J. Bacteriol* **170**:528-33; OmpF (Pages *et al. Biochemimie* (1990) **72**:169-76); PhoE (van der Ley *et al. J. Biol. Chem.* **261**:12222-5); pilin (So *et al. Curr Top in Microbiol & Immunol* (1985) **118**:13-28); pldA (de Geus *et al. EMBO J.* (1984) **3**(8): 1799-1802); BtuB, FepA, FhuA, IutA, FecA, and FhuE (Gudmundsdottir et al, (1989) *J. Bacteriol* **171**(12):6526-33); GIP-

anchored protein INP(Kim et al. (1999) *Lett Appl Microbiol* **29**(5):292-297) and β-autotransporter protein AIDA (Veiga et al. (1999) *Mol Microbiol* **33**: 1232-1243), and other outer membrane lipoproteins such as TratT, Pal, Oprl, OsmB, N1pB and BlaZ, . Numerous coat proteins residing on the surface of bacterial spores have also been identified. Their corresponding gene sequences have subsequently been isolated. For example, Donovan et al. reported the identification of *Bacillus subtilis* spore coat CotD and CotC genes (Donovan et al. (1987) *J. Mol. Biol.* **196**:1-10). Characterization of these and other surface proteins are detailed in Pierre Cornelis et al. (2000) *Curr. Opin. Biotech.* **11**(5):450-454; Lang et al. (2000) *Int. J. Med. Microbiol.* **290**: 579-585; Daugherty et al. (1999) *Protein Engineering* **12** (7): 613-621; U.S. Patent Nos. 5,837,500 and 5,348,867 as well as the references cited therein.

148 The signal peptide and the membrane translocation domain of these and other bacterial outer-surface proteins are well known in the art. The signal peptide generally consists of the first 5 to 30 N-terminal amino acids of the protein. The membrane translocation domain typically comprises one or more membrane spanning segments that are readily identifiable via computer-assisted conventional sequence analyses. One of ordinary skill in the art can readily obtain the appropriate polypeptide and nucleotide sequences using conventional synthetic and recombinant technology. Where desired, the signal peptide of one outer-surface protein may be attached in-frame to the membrane translocation domain of another outer-surface protein, or vice versa. It has been shown that such a chimera can be expressed on the bacterial outer surface (U.S. Patent No. 5,837,500). As such, a signal peptide leader peptide of any one of the above-mentioned bacterial outer-surface proteins can be linked in-frame to the membrane translocation domain of suitable length of any native bacterial outer-surface proteins. Similarly, the translocation domain of any one of the aforementioned outer-surface

proteins can be fused in-frame to a signal peptide of any protein of bacterial or other origins, known to be capable of directing the fusion to the bacterial periplasm.

149 The bacterial expression vector of the present invention lacks sequences encoding any functional bacterial outer-surface proteins. Any of the aforementioned outer-surface sequences are candidate sequences to be excluded while constructing the subject expression vector. As used herein, the term "functional" is meant that the encoded outer-surface proteins retain the ability to facilitate or direct the genetic package to assemble the polypeptide of interest onto its outer surface. The loss of the "function" may be attributed to modification(s) resulting in (1) loss of a functional signal peptide that direct the intracellular translocation of the outer-surface protein into the [periplasm] of the bacterial cells, where the signal peptide is then cleaved off; (2) loss of function of the membrane translocation domain that translocates the mature polypeptide onto the bacterial cell membrane; and/or (3) introduction of internal stop codons to prevent expression of any functional outer-surface proteins.

150 The two adapter sequences linked to the expression and helper vectors have the same structural and functional characteristics as the ones employed in the subject phage display system. Any adapter sequence applicable for constructing the phage display system is equally suited for generating the bacterial display system. Thus, the criteria and procedures for selecting and preparing the pairing adapters are not repeated in this section.

151 Suitable bacterial genetic packages include all bacterial strains, which can be grown in culture and can be engineered to display exogenous polypeptide on their outer surface, and are compatible with affinity selection. Preferred genetic packages are gram-negative bacteria. Non-limiting examples of preferred species include *Salmonella typhimurium*, *Bacillus subtilis*, *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Klebsiella*

pneumonia, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Bacteroides nodosus*, *Moraxella bovis*, and especially *Escherichia coli*.

152 Bacterial spores have desirable properties as genetic packages. Spores are much more resistant than vegetative bacterial cells to chemical and physical agents, and hence permit the use of a great variety of test conditions. For instance, bacteria of the genus *Bacillus* form endospores that are extremely resistant to damage by heat, radiation, desiccation, and toxic chemicals (reviewed by Losick et al. *Ann. Rev. Genet.* (1986) **20**:625-669). In addition, the Bacillus spores neither actively metabolize nor alter the proteins on their surface. This phenomenon is attributed to extensive intermolecular crosslinking of the coat proteins. Other spores useful as genetic packages are exospores, such as spores of Streptomyces.

Proof - 96500

Other Consideration for Constructing the Subject Phage and Bacterial Display Systems:

153 The vectors of the present invention generally comprise transcriptional or translational control sequences required for expressing the exogenous polypeptide. Suitable transcription or translational control sequences include but are not limited to replication origin, promoter, enhancer, repressor binding regions, transcription initiation sites, ribosome binding sites, translation initiation sites, and termination sites for transcription and translation.

154 The origin of replication (generally referred to as an ori sequence) permits replication of the vector in a suitable host cell. The choice of ori will depend on the type of host cells and/or genetic packages that are employed. Where the host cells are prokaryotes and the genetic packages are phage particles, the expression vector typically comprises two ori sequences, one directing autonomous replication of the vector within the prokaryotic cells, and the other ori supports packaging of the phage particles.

Preferred prokaryotic ori is capable of directing vector replication in bacterial cells. Non-limiting examples of this class of ori include pMB1, pUC, as well as other *E. Coli* origins. Preferred ori supporting packaging of the phage particles includes but is not limited to f1 ori, Pf3 phage replication ori.

155 As used herein, a "promoter" is a DNA region capable under certain conditions of binding RNA polymerase and initiating transcription of a coding region located downstream (in the 3' direction) from the promoter. It can be constitutive or inducible. In general, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence is a transcription initiation site, as well as protein binding domains responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes.

156 The choice of promoters will largely depend on the host cells in which the vector is introduced. For prokaryotic cells, a variety of robust promoters are known in the art. Preferred promoters are lac promoter, Trc promoter, T7 promoter and pBAD promoter.

157 Suitable promoter sequences for other eukaryotic cells include the promoters for 3-phosphoglycerate kinase, or other glycolytic enzymes, such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. Other promoters, which have the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization.

158 In constructing the subject vectors, the termination sequences associated with the exogenous sequence are also inserted into the 3' end of the sequence desired to be transcribed to provide polyadenylation of the mRNA and/or transcriptional termination signal. The terminator sequence preferably contains one or more transcriptional termination sequences (such as polyadenylation sequences) and may also be lengthened by the inclusion of additional DNA sequence so as to further disrupt transcriptional read-through. Preferred terminator sequences (or termination sites) of the present invention have a gene that is followed by a transcription termination sequence, either its own termination sequence or a heterologous termination sequence. Examples of such termination sequences include stop codons coupled to various polyadenylation sequences that are known in the art, widely available, and exemplified below. Where the terminator comprises a gene, it can be advantageous to use a gene which encodes a detectable or selectable marker; thereby providing a means by which the presence and/or absence of the terminator sequence (and therefore the corresponding inactivation and/or activation of the transcription unit) can be detected and/or selected.

159 In addition to the above-described elements, the vectors may contain a selectable marker (for example, a gene encoding a protein necessary for the survival or growth of a host cell transformed with the vector), although such a marker gene can be carried on another polynucleotide sequence co-introduced into the host cell. Only those host cells into which a selectable gene has been introduced will survive and/or grow under selective conditions. Typical selection genes encode protein(s) that (a) confer resistance to antibiotics or other toxins, e.g., ampicillin, kanamycin, neomycin, G418, methotrexate, etc.; (b) complement auxotrophic deficiencies; or (c) supply critical nutrients not available from complex media. The choice of the proper marker gene will depend on the host cell, and appropriate genes for different hosts are known in the art.

160 In a preferred embodiment, the vector is a shuttle vector, capable of replicating in at least two unrelated expression systems. In order to facilitate such replication, the vector generally contains at least two origins of replication, one effective in each expression system. Typically, shuttle vectors are capable of replicating in a eukaryotic expression system and a prokaryotic expression system. This enables detection of protein expression in the eukaryotic host (the expression cell type) and amplification of the vector in the prokaryotic host (the amplification cell type). Preferably, one origin of replication is derived from SV40 and one is derived from pBR322 although any suitable origin known in the art may be used provided it directs replication of the vector. Where the vector is a shuttle vector, the vector preferably contains at least two selectable markers, one for the expression cell type and one for the amplification cell type. Any selectable marker known in the art or those described herein may be used provided it functions in the expression system being utilized

161 The vectors embodied in this invention can be obtained using recombinant cloning methods and/or by chemical synthesis. A vast number of recombinant cloning techniques such as PCR, restriction endonuclease digestion and ligation are well known in the art, and need not be described in detail herein. One of skill in the art can also use the sequence data provided herein or that in the public or proprietary databases to obtain a desired vector by any synthetic means available in the art. Additionally, using well-known restriction and ligation techniques, appropriate sequences can be excised from various DNA sources and integrated in operative relationship with the exogenous sequences to be expressed in accordance with the present invention.

162 The exogenous sequences expressed by the subject display systems can be heterolgous sequences of any length. "Heterologous" means derived from a genetically distinct entity from the rest of the entity to which it is being compared. For instance, the heterologous sequence can be a gene not normally expressed in the genetic package (e.g.

bacterial cell or phage particle). Alternatively, the heterologous sequence can be a gene native to the genetic package but is linked to a coding sequence other than the native sequence that the gene is naturally operably linked to. Furthermore, the heterologous sequence may encode random or predetermined polypeptide.

163 The exogenous sequence expressed by the subject systems also can be characterized based on one or more of the following features: species origin, developmental origin, primary structural similarity, involvement in a particular biological process, association with or resistance to a particular disease or disease stage, tissue, sub-tissue or cell-specific expression pattern, and subcellular location of the expressed gene product.

164 In one aspect, the exogenous sequence may be any sequence expressed in an entity other than the genetic package, such as a plant cell, animal cell or a yeast cell.

165 In another aspect, the exogenous sequences are of a specific developmental origin, such as those expressed in an embryo or an adult organism, during ectoderm, mesoderm, or endoderm formation in a multi-cellular animal, or during development of leaves, tubers, bud of a plant.

166 In yet another aspect, the exogenous sequences belong to a family of genes, or a sub-family of genes that share primary structural similarities. Structural similarities can be discerned with the aid of computer software described above. Non-limiting examples of gene families include those encoding proteinase, proteinase inhibitors, cell surface receptors, protein kinases (e.g. tyrosine, serine/threonine or histidine kinases), trimeric G-proteins, cytokines, PH-, SH2-, SH3-, PDZ-domain containing proteins, and any of those gene families published by the Institute for Genomic Research (TIGR), Incyte Pharmaceuticals, Inc., Human Genome Sciences Inc., Monsanto, and Celera.

167 In yet another aspect, the exogenous sequences are involved in a specific biological process, including but not limited to cell cycle regulation, cell differentiation,

chemotaxsis, apoptosis, cell motility and cytoskeletal rearrangement. In still another aspect, the exogenous sequences embodied in the invention are associated with a particular disease or with a specific disease stage. Such sequences include but are not limited to those associated with autoimmune diseases, obesity, hypertension, diabetes, neuronal and/or muscular degenerative diseases, cardiac diseases, endocrine disorders, any combinations thereof.

168 In yet still another aspect, the exogenous sequences encompass those exhibiting restricted expression patterns. Non-limiting exemplary gene transcripts of this class include those that are not ubiquitously expressed, but rather are differentially expressed in one or more of the plant tissues including leaf, seed, tuber, stems, root, and bud; or expressed in animal body tissues including heart, liver, prostate, lung, kidney, bone marrow, blood, skin, bladder, brain, muscles, nerves, and selected tissues that are affected by various types of cancer (malignant or non-metastatic), affected by cystic fibrosis or polycystic kidney disease. Additional examples of non-ubiquitously expressed sequences are those whose protein products are localized to certain subcellular locations: extracellular matrix, nucleus, cytoplasm, cytoskeleton, plasma and/or intracellular membranous structures which include but are not limited to coated pits, Golgi apparatus, endoplasmic reticulum, endosome, lysosome, and mitochondria.

169 The subject display system may comprise a selectable library of genetic packages. The packages may express the same or distinct exogenous sequences. In one aspect, the library of genetic packages encodes a population of random or predetermined polypeptides. In another aspect, the library encodes a population of cDNAs that are derived from cells of specific host origin, tissue origin, developmental stage, or particular disease state.

170 A particularly preferred library encodes a population of antigen-binding units. The antigen-binding units may be monomeric or multimeric. Monomeric antigen-

binding units are commonly referred as single-chain antigen-binding units (Sc Abus), whereas the multimeric antigen-binding units are referred to herein as non-single-chain antigen-binding units (Nsc Abus).

171 The Nsc Abus that can be displayed by the subject system may be either "monovalent" or "multivalent." The displayed multivalent Abus can be further characterized as "monospecific" or "multispecific" Abus. To display multimeric Abus, two sets of expression vectors, one comprising the light chain (L) variable regions, and the other comprising the heavy chain (H) variable regions must be employed. Whereas the expressed antibody regions dimerize through preferably heterodimerization sequences fused in-frame with the antibody regions, one of the expressed antibody regions must comprise additionally, an adapter capable of pairwise interaction with the other adapter provided by the helper vector.

172 Nucleotide sequences corresponding to various regions of L or H chains of an existing antibody can be readily obtained and sequenced using convention techniques including but not limited to hybridization, PCR, and DNA sequencing. Hybridoma cells that produce monoclonal antibodies serve as a preferred source of antibody nucleotide sequences. A vast number of hybridoma cells producing an array of monoclonal antibodies may be obtained from public or private repositories. The largest depository agent is American Type Culture Collection (http://www.atcc.org), which offers a diverse collection of well-characterized hybridoma cell lines. Alternatively, antibody nucleotides can be obtained from immunized or non-immunized rodents or humans, and form organs such as spleen and peripheral blood lymphocytes. Specific techniques applicable for extracting and synthesizing antibody nucleotides are described in Orlandi *et al.*(1989) *Proc. Natl. Acad. Sci. U.S.A* **86**: 3833-3837; Larrick *et al.* (1989) *Biochem. Biophys. Res. Commun.* **160**:1250-1255; Sastry *et al.* (1989) *Proc. Natl. Acad. Sci., U.S.A.* **86**: 5728-5732; and U.S Patent No. 5,969,108.

173 The antibody nucleotide sequences may also be modified, for example, by substituting the coding sequence for human heavy and light chain constant regions in place of the homologous non-human sequences, or vice versa. In that manner, chimeric antibodies are prepared that retain the binding specificity of the original antibody.

174 Where desired, the exogenous sequence may comprise sequences coding for moieties that facilitate detection of the expression and purification of the protein product. Examples of such moieties are known in the art and include those encoding reporter proteins such as β-galactosidase, β -lactamase, chloramphenicol acetyltransferase (CAT), luciferase, green fluorescent protein (GFP) and their derivatives. Other sequences that facilitate purification may code for epitopes such as Myc, HA (derived from influenza virus hemagglutinin), His-6, FLAG, or the Fc portion of immunoglobulin, glutathione S-transferase (GST), and maltose-binding protein (MBP).

Host Cells of the Present Invention:

175 The invention provides host cells comprising the expression and/or helper vectors described above. The expression vectors can be introduced into a suitable prokaryotic or eukaryotic host cell by any of a number of appropriate means, including electroporation, microprojectile bombardment; lipofection, infection (where the vector is coupled to an infectious agent), transfection employing calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances. Depending on the features of the host cells, one of ordinary skill can readily practice one or more of the appropriate means that are well established in the art.

176 Once introduced into a suitable host cell, expression of the exogenous sequence can be determined using any nucleic acid or protein assay known in the art. For example, the presence of transcribed mRNA of the exogenous sequence can be detected

and/or quantified by conventional hybridization assays (e.g. Northern blot analysis), amplification procedures (e.g. RT-PCR), SAGE (U.S. Patent No. 5,695,937), and array-based technologies (see e.g. U.S. Pat. Nos. 5,405,783, 5,412,087 and 5,445,934), using probes complementary to any region of the exogenous sequence.

177 Expression of the exogenous sequence can also be determined by examining the expressed protein product. A variety of techniques are available in the art for protein analysis. They include but are not limited to radioimmunoassays, ELISA (enzyme linked immunoradiometric assays), "sandwich" immunoassays, immunoradiometric assays, in situ immunoassays (using *e.g.*, colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunoflourescent assays, and PAGE-SDS.

178 The host cells of this invention can be used, *inter alia*, as repositories of the subject exogenous sequences, vectors, or as vehicles for producing and screening desired polypeptides such as Abus based on their binding specificities.

Uses of the Adapter-Directed Display Systems of the Present Invention:

179 The adapter-directed display systems of this invention have several specific uses. First, the systems permit the production of soluble monomeric and multimeric exogenous polypeptides in suitable host cells. Second, the systems allow the display of monomeric and multimeric polypeptides on selected genetic packages. The subject display systems also can be used to create libraries of random or predetermined polypeptides, full-length proteins, and protein domains for a variety of purposes. For instance, the displayed libraries can be employed for mapping epitopes and mimotopes, identifying antagonists and agonists of various target proteins, engineering antibodies, optimizing antibody specificities and creating novel binding activities.

180 Accordingly, the present invention provides a method of detecting the presence of a specific interaction between a test agent and an exogenous polypeptide that is displayed on a genetic package. The method involves the steps of: (a) providing a genetic package of the subject display system that presents the exogenous polypeptide; (b) contacting the genetic package with the test agent under conditions suitable to produce a stable polypeptide-agent complex; and (c) detecting the formation of the stable polypeptide-agent complex on the genetic package, thereby detecting the presence of the specific interaction.

181 For the purposes of this invention, a "test agent" is intended to include, but not be limited to a biological or chemical compound such as a simple or complex organic or inorganic molecule, a protein, carbohydrate, lipid, polynucleotide or combinations thereof. A vast array of compounds can be synthesized, for example oligomers, such as oligopeptides and oligonucleotides, and synthetic organic compounds based on various core structures, and these are also included in the term "agent." In addition, various natural sources can provide compounds for screening, such as plant or animal extracts, and the like. It should be understood, although not always explicitly stated that the agent is used alone or in combination with another agent, having the same or different biological activity as the agents identified by the inventive screen. Preferred agents are candidate diagnostics and/or therapeutics, such as those capable of modulating the signal transduction pathways of a cell.

182 In a separate embodiment, the present invention provides a method of obtaining a polypeptide with desired property. The method comprises the steps of (a) providing a selectable library of the subject display system; and (b) screening the selectable library to obtain at least one genetic package displaying a polypeptide with the desired property. The method may further comprise the step of isolating the genetic package that displays a polypeptide having the desired property. Such isolation of the genetic

package may involve obtaining a nucleotide sequence from the genetic package that encodes the desired polypeptide. The desired property encompasses the ability of the polypeptide to specifically bind to an agent of interest. The selected polypeptide with the desired property may fall within one or more classes of the following molecules, namely antigen-binding unit, cell surface receptor, receptor ligand, cytosolic protein, secreted protein, nuclear protein, and functional motif thereof. The choice of specific agent to be tested and the libraries of exogenous polypeptides to be displayed will depend on the intended purpose of the screening assay.

Isolating antibodies exhibiting desired binding specificity or affinity:

183 One of the most powerful applications of phage and bacterial display is in the arena of antibody engineering. It has been shown that scFv antigen-binding units can be expressed on the surface of both phage particles and bacterial cells with no apparent loss of binding specificity and affinity (McCafferty et al. (1990) *Nature* **348**:552-554; Daugherty et al. (1999) *Protein Engineering* **12** (7): 613-621). It has also been demonstrated that functional Nsc Abus such as Fab fragments can be expressed the on phage surface. Today, antibodies to many diverse antigens have been successfully isolated using phage display technology.

184 The subject phage display system is particularly suited for this application because the system allows presentation of a vast diverse repertoire of Abus. In many respects the subject phage display system mimics the natural immune system. Antigen-driven stimulation of Abu can be achieved by selecting for high-affinity binders from a phage display library of Abus. The large number of chain permutations that occur during recombination of H and L chain genes in developing B cells can be mimicked by shuffling the cloned H and L chains as DNA, and protein and through the use of site-specific recombination (Geoffory et al. (1994) *Gene* **151**:109-113). The somatic

mutation can also be matched by the introduction of mutations in the CDR regions of phage-displayed Abus.

185 The Abus with desired binding specificity or affinity can be identified using a form of affinity selection known as "panning" (Parmley and Smith (1988) *Gene* **73**:305-318). The library of Abus is first incubated with an antigen of interest followed by the capture of the antigen with bound phage. The phage recovered in this manner can then be amplified and again gain selected for binding to the antigen, thus enriching for those phages that bind the antigen of interest. Usually, three to four rounds of selection can be accomplished with a week, leading to the isolation of one to hundreds of binding phages. Thus rare phage expressing desired Abu can easily be selected from greater than 10^8 different individuals in one experiment. The primary structure of the binding Abu is then deduced by nucleotide sequence of the individual phage clone. When human V_H and V_L regions are employed in the displayed Abus, the subject display systems allow selection of human antibodies without further manipulation of a non-human Abu.

Generating novel proteins including Abus with improved binding specificity or affinity:

186 Using the subject display systems, one can obtain a replicable genetic package that displays a polypeptide, such as an Abu, having high affinity and specificity for a target protein. Such a package carries both amino acids of the binding polypeptide and a polynucleotide encoding the binding product. The presence of the polynucleotide facilitates recombinant expression and subsequent manipulation of the binding protein. For instance, the polynucleotide coding for the binding protein can be mutagenized by cassette mutagenesis, error-prone PCR, or shuffling to generate a refined repertoire of altered sequences that resemble the parent polynucleotide. Upon screening the refined

repertoire of novel binding proteins, those exhibiting improved binding specificity or affinity can be identified.

Mapping antigenic epitopes:

187 Traditionally, epitope mapping of an antigen has relied heavily on physical chemical analysis. These approaches have included: (1) fragmenting the purified antigen with various proteases, identifying reactive fragments, and sequencing them; (2) chemical modification experiments in which residues interaction with the antigen-binding unit are protected from modification; (3) synthesizing a series of peptides corresponding to the primary structure of the antigen; and (4) direct physical characterization using NMR or X-ray crystallography. All of these methods are labor intensive and generally not amenable to high-throughput analyses. Phage or bacterial display provides a highly efficient and robust alternative for localizing the antigenic epitope. Fragments of DNA that encode portions of the antigen can be expressed as the exogenous polypeptides by the subject expression vectors. The genetic packages (e.g. phage, bacterial cells or spores) can then be tested with the antibody to determine which displayed fragments react with the antibody. This application of display technology has been widely used in the art and shown to be successful for determining the antigenic epitopes of a variety of molecules.

Mapping binding epitopes of monoclonal and polyclonal Abus

188 The subject display system also can be used to present random peptide libraries for mapping the specificity of the antigen-binding sites. Random peptide libraries represent a source of sequences from which epitopes and mimotopes can be operationally defined. With such a library, one can identity and obtain peptide

competitors for antigen-antibody interactions, and thus map accessible and/or functional sites of numerous Abus.

Identifying ligands of receptors and other modulators of signal transduction pathway:

189 The subject display systems can also be employed to identify ligands for receptors. The process generally proceeds with subjecting a population of genetic packages expressing the test ligands to the receptors, followed by identifying the packages bound to the receptors. Alternatively, the receptors may be presented in the genetic packages. Those that are bound to the test ligands are then isolated. For identification of peptide ligands, random peptide libraries are preferred staring materials for performing the assay. The same approach is applicable for identifying other modulators of the signal transduction pathways of a cell.

190 The activity of cells is regulated by external signals that stimulate or inhibit intracellular events. The process by which stimulatory or inhibitory signals are transmitted into and within a cell to elicit an intracellular response is referred to as signal transduction. Proper signal transduction is essential for proper cellular function. Over the past decades, numerous cellular signaling molecules have been identified, cloned and characterized. Non-limiting examples of the signaling proteins include cell surface receptors, protein kinases (e.g. tyrosine, serine/threonine or histidine kinases), trimeric G-proteins, cytokines, SH2-, SH3-, PH-, PDZ-, death-domain containing proteins, and any of those gene or protein families published by Human Genome Sciences Inc., Celera, the Institute for Genomic Research (TIGR), and Incyte Pharmaceuticals, Inc. Cascades of signal transduction events mediated by the ever-growing families of signaling proteins have been elucidated and found to play a central role in a variety of biological responses. Among them are cell cycle regulation, cell differentiation, apoptosis, chemotaxsis, cell motility and cytoskeletal rearrangement

(Cantley et al. (1991) *Cell* **64**:281-302); Liscovitch et al. (1994) *Cell* **77**:329-334). Defects in various components of signal transduction pathways have also been found to account for a vast number of diseases, including numerous forms of cancer, vascular diseases and neuronal diseases. Indeed, agents capable of modulating signaling pathways (i.e. modulators of the signal transduction pathway) have long been acknowledged as potential diagnostic and/or therapeutic agents.

191 The modulators of the present invention is characterized by their ability to (1) bind to intracellular signaling proteins presented on the subject genetic packages; or (2) compete for binding to the displayed signaling proteins in the presence cellular proteins that are normally associated with the signaling proteins. The modulators can be an agonist or an antagonist of a target signal protein.

Express Mail Label No. EV339905014US

Expressing cDNA libraries:

192 The subject display systems are particularly suited for expressing cDNA libraries. As noted above, the previously reported fusion systems including the gene III and gene VIII phage display systems restrict the point of insertion to the 5' end of the outer-surface sequence. The exogenous polypeptide thus must be linked to the N-terminus of the outer-surface proteins. Consequently, cDNA libraries containing fragments of coding sequences of all reading frames cannot be fully expressed by these fusion systems due to the disruption of reading frames by internal stop codons. The subject systems, however, do not suffer from this drawback of unidirectional cloning because the exogenous sequence is not fused in frame with the outer-surface sequence.

193 cDNA display can be a useful technique for defining protein-protein interactions. Expression and screening of cDNA libraries greatly facilitate the identification of novel genes based on the ability of the expressed product to bind a known protein of particular interest. The cDNA-encoded proteins can be expressed on the surface of the subject

genetic packages, which can then be tested against a particular immobilized target *in vitro* via biopanning enrichment as detailed above.

194 The ability of a displayed exogenous polypeptide or a library of random or predetermined polypeptide to specifically bind to a test agent can be tested by a variety of procedures well established in the art. Generally selection is preferably performed using affinity chromatography. The method typically proceeds with binding the genetic packages a test-agent-coated plates, column matrices, cells or to biotinylated agents in solution followed by capture. The genetic packages bound to the solid phase are washed and then eluted by soluble hapten, acid or alkali. Alternatively, increasing concentrations of the test agent can be used to dissociate the genetic packages from the affinity matrix. For certain Abus with extremely high affinity or avidity to the test antigen, efficient elution may require high pH or mild reducing solution as described in WO 92/01047.

195 To avoid potential difficulties in recovering the bound polypeptide with the desired binding specificities, protease cleavage sites may be introduced between the adapter and exogenous polypeptide. Cleavage sites applicable for this purpose include but are not limited to Factor X, trypsin, and thrombin recognition sites. After binding the genetic packages to an affinity matrix and washing the non-specific packages, the remaining packages that display the exogenous polypeptide with the desired affinity can be collected by washing the antigen-affinity matrix with protease under conditions suitable for digestion at the cleavage site. Such digestion would release the exogenous polypeptide from the genetic packages such as phage particles.

196 An alternative procedure to the above is to take the affinity matrix which has retained the strongly bound phage or bacterial particles and extract their nucleic acids, for example by boiling in SDS solution. Extracted nucleic acids can be used to directly

transform E. coli host cells or alternatively the exogenous sequence can be amplified by PCR using suitable primers.

197 The efficiency of selection is likely to depend on a combination of several factors, including the kinetics of dissociation during washing, and whether multiple copies of the exogenous polypeptides on a single phage or bacterium can simultaneously bind to the test agent on a solid support. For example, antibodies with fast dissociation kinetics (and weak binding affinities) should be retained by use of short washes, multivalent display and a high coating density of antigen at the solid support. Conversely, the selection of Abus with slow dissociation kinetics (and good binding affinities) should be favored by use of long washes, monovalent phages, and a low coating density of antigen.

198 Alternatively, specific binding to a given agent can be assessed by cell sorting. The technique involves presenting the exogenous polypeptide on genetic packages such as phage particles that are adhered to host cells to be sorted, then labeling the target cells with test agents that are coupled to detectable moieties, followed by separating the labeled cells from the unlabeled ones in a cell sorter. A sophisticated cell separation method is fluorescence-activated cell sorting (FACS). Cells traveling in single file in a fine stream are passed through a laser beam, and the fluorescence of each cell bound by the fluorescent label is then measured.

199 Where desired, the repertoire of exogenous polypeptides can be pre-selected against an unrelated test agent to counter-select the undesired polypeptide. For instance, a repertoire of Abus may be counter-selected against a unrelated antigen. The repertoire may also be pre-selected against a related agent in order to isolate, for example, anti-idiotypic Abus. The subject display systems enables rapid isolation of Abus with desired specificities. Many of the isolated Abus would be expected to be difficult or impossible to obtain through conventional hybridoma or transgenic animal technology.

200 Subsequent analysis of the eluted Abus may involve protein sequencing for delineating the amino acid sequences of the L and H chains. Based on the deduced amino acid sequences, the cDNA encoding the antibody polypeptides can then be obtained by recombinant cloning methods including PCR, library screening, homology searches in existing nucleic acid databases, or any combination thereof. Commonly employed databases include but are not limited to GenBank, EMBL, DDBJ, PDB, SWISS-PROT, EST, STS, GSS, and HTGS.

Kits comprising the vectors of the present invention

201 The present invention also encompasses kits containing the expression and helper vectors of this invention in suitable packaging.

202 Each kit necessarily comprises the reagents which render the delivery of vectors into a host cell possible. The selection of reagents that facilitate delivery of the vectors may vary depending on the particular transfection or infection method used. The kits may also contain reagents useful for generating labeled polynucleotide probes or proteinaceous probes for detection of exogenous sequences and the protein product. Each reagent can be supplied in a solid form or dissolved/suspended in a liquid buffer suitable for inventory storage, and later for exchange or addition into the reaction medium when the experiment is performed. Suitable packaging is provided. The kit can optionally provide additional components that are useful in the procedure. These optional components include, but are not limited to, buffers, capture reagents, developing reagents, labels, reacting surfaces, means for detection, control samples, instructions, and interpretive information.

203 Further illustration of the development and use of subject display systems, host cells, and genetic packages are provided in the Example section below. The examples

are provided as a guide to a practitioner of ordinary skill in the art, and are not meant to be limiting in any way.

Express Mail Label No. EV 339 908 425 US

EXAMPLES

Example 1: Preparation and uses of KO7kpn helper phage

A. Construction of KO7kpn vector:

204 The KO7kpn vector was constructed by modifying a well-characterized vector, namely M13KO7 (from Amersham Pharmacia) according to the procedure detailed below. The resulting vector is identical to KO7 except that a unique KpnI restriction site has been inserted into the gene III leader sequence without disrupting the gene III coding region (see Figures 3A and 3B).

205 The KpnI site was introduced into the gene III leader sequence of KO7 helper phage vector by PCR- based site-directed mutagenesis. The KO7 genome was amplified by PCR using the following primers which contain KpnI sites: p3KN1: 5'-TTTAGTGGTA CCTTTCTATTCTCACTCCGCTG-3' and p3KN2: 5'-TAGAAAGGTACCACTAAAG GAATTGCGAATAA-3'. These primers share partial sequence homology to gene III leader sequence.

206 PCR was performed in a 100 ul reaction mixture containing 100 ng KO7 vector DNA, 20 pmol each of primers, 250 uM dNTP, and 1X pfu buffer and pfu DNA polymerase (Stratagene). The reaction mixture was initially incubated at about 96 ^{0}C and then subjected to 15 cycles of PCR in a thermocycler as follows:

denaturation 96°C, 30 seconds
annealing 55°C, 30 seconds
extension 72°C, 10 minutes

207 After amplification, the products were gel purified, cut with KpnI and ligated to transform TG1 bacterial cells by electroporation. The bacterial cells were sleeted for kanamycin resistance. Specifically, the kanamycin-resistant (Kan^{R}) colonies were grown in 96-well microtiter plates in 2xYT medium with 70 ug/ml Kanamycin, and

supernatants were used for phage screening by phage ELISA assay to eliminate the loss-of-function mutants caused by PCR errors. Briefly, the phage ELISA was conducted as follows: A 100 ul of the supernatant containing phage particles was employed to coat wells of the ELISA plates at 4 °C overnight. After blocking with 5% milk in PBS buffer for 30 minutes at room temperature, the phage particles bound on ELISA plates were further incubated with 100 ul of HRP-conjugated anti-M13 antibody (Amersham Pharmcia) for 1 hour at room temperature. The free anti-M13 antibodies were washed away by PBS containing 0.05% Tween 20. The substrate ABTS [2,2'Azino-bis(3-ethylbenzthiazoline-6-sulfonic acid)] was then added. The HRP activity was determined by the absorbance at 405 nm. Figure 2 shows the results of a phage ELISA screen for kanamycin-resistant, phage-positive clones. 48 clones were screened for phage generation. The clones C2, B3, B7, B9, A12 were phage positive. The DNAs extracted from clones B7, B9 and A12 were prepared from TG1 cultures. Double digestion of vector DNA with Acc65I (isoschizomer of KpnI) and BamHI showed a 600 bp DNA fragment, which confirms the presence of KpnI site in all of the three KO7kpn vector clones.

B. KO7kpn phage generation:

208 The Kan^R TG1 supernatant containing KO7kpn helper phages produced from B9 clone was streaked on a 2x YT agar plate. 4 ml of soft agar mixed with 0.5 ml of TG1 culture (OD600 = 0.5) was poured on the plate. Phage plaques were formed after incubation at 37 °C overnight. A single phage plaque was picked and used to inoculate 10 ml 2x YT culture with 70μg/ml kanamycin. After incubating at 37° C for 2 hours with constant shaking at 250 rpm, the culture was transferred to a 2 liter flask containing 500 ml 2x YT with 70 μg/ml kanamycin. The culture was incubated overnight with

constant shaking. The phages in the supernatants were then precipitated using polyethylene glycol (PEG)/NaCl, and re-suspended in phosphate-buffed saline (PBS). The phage concentration was determined by OD_{268} measurement. Generally, a reading of 1 unit at OD_{268} indicates that the supernatant contains approximately 5 X 10^{12} phage/ml. The recorded phage yield for KO7kpn helper phage was approximately 1-2 X 10^{12}/ml, which was very similar to that of the M13KO7 helper phages.

C. Use of KO7Kpn helper phage for phage display:

209 The coding sequence of scFv antibody AM2 was subcloned into phage display vector pABMD1 (Figure 22) that expresses scFv-pIII fusions. TG1 cells carrying the display vectors were grown to OD600 = 0.6, and superinfected by KO7kpn helper phage at MOI=10. The infected TG1 cells were grown in 2xYT/Amp/Kan at 30 °C overnight. The phagemid particles were precipitated twice by PEG/NaCl from culture supernatants, and resuspended in PBS. The scFv-pIII fusions displayed on phage were detected via phage ELISA assay. Briefly, 0.2 ug AM2-antigens were first coated onto 96-well ELISA plates at 4 °C overnight. After 5% milk/PBS blocking, the phage solution in 2% Milk/PBS was placed onto the ELISA plates for 1 hour. The phage bound to antigen was detected by incubation with HRP-conjugated anti-M13 antibody. The substrate ABTS [2,2'Azino-bis(3-ethylbenzthiazoline-6-sulfonic acid)] was used for measurement of HRP activity. The phagemids generated from the TG1 cells carrying pABMD1-AM2 / KO7kpn vectors showed a very strong antigen binding activity, indicating the functional display of scFv antibody AM2 by KO7kpn helper phage. This series of experiments serve as positive controls for reagents employed for constructing and using the subject display systems.

Example 2: Preparation of an Adapter-Directed Display System Comprising GM-UltraHelper phage vectors

A. Construction of GM-UltraHelper phage vector:

210 The pABMC6 vector was constructed by replacing the sequence between the XbaI and BglII sites of vector pABMD1 (Figure 22A) with a synthetic DNA fragment encoding a partial gene III leader sequence with a KpnI site, and coding sequences for GR2 domain (the coiled-coil domain of the human $GABA_B$ receptor 2) and Myc-tag (Figure 4). The sequence for GR2-Myc domain was directly fused with the pIII coding sequence in pABMD1 vector, and was confirmed by DNA sequencing.

211 The GM-UltraHelper phage vector was constructed by replacing the KpnI/BamHI fragment encoding a partial pIII leader (amino acid residues 11-19) and partial pIII protein (amino acid residues 1-197) in the KO7Kpn helper vector with the corresponding fragment encoding a partial pIII leader and the adaptor2-pIII fusion protein from the pABMC6 vector. The resulting GM-UltraHelper phage vector (Figure 5) comprises an engineered gene III fusion in which a GR2 domain and a Myc-tag sequence (for detection of engineered pIII protein) are fused in-frame with gene III (Figure 5B).

212 The B9 KO7kpn helper phage clone (see Example 1) was used for constructing the GM-UltraHelper phage vector. After subcloning of the engineered gene III fragment into KO7kpn phage vector, 20 kanamycin-resistant colonies were grown in 96-well microtiter plates in 2xYT medium with 70 ug/ml Kanamycin. The supernatant was used for phage screening by phage ELISA assay as described in Example 1. 19 clones were able to generate phage particles.

213 To confirm that GM-UltraHelper phages were packaged with GR2-Myc-pIII fusion proteins, western blot using anti-Myc antibody (9E10 from BD Pharmingen) was carried out to detect the engineered pIII fusion. Briefly, 1-4 x 10^{11} Phage particles from

four clones (clones 6, 9, 18, and 20) were heated for 10 min in SDS sample buffer (2% SDS, 5% β-mercaptoethnol, 10% glycerol, 0.67 M Trice-HCl, pH 6.8). The denature sample was subjected to SDS-PAGE. The proteins in the SDS gel was then transferred to PVDF membrane which was subsequently probed with 2 ug/ml 9E10 antibody in 5% milk/PBS. Myc-tagged proteins were detected by anti-mouse antibody-AP conjugate and BCIP/NBT AP substrate (Sigma). As shown in Figure 6, a single protein band was detected in all four clones by the anti-Myc antibody. No Myc-containing band was detected in the negative control M13KO7 helper phage. This experiment demonstrates that GR2-Myc-pIII fusion proteins were assembled into UltraHelper phage particles. The assembly of GR2-Myc-pIII fusions was further confirmed by ELISA assays (see Figure 7).

Express Mail No. EV 339933705 US

214 It has been well known that M13 phage is resistant to trypsin. The screening for protease cleavage sites revealed that there are seven trypsin cleavage sites in the GR2-Myc domain of the GM-UltraHelper phage (Figure 5C). To test whether the GR2-Myc domain can be cleaved from the phage surface by trypsin, GM-Ultrahelper phages from clone 18 were exposed to different concentrations of trypsin for 30 minutes at 37° C, and then trypsin inhibitor was added to stop the reaction. Figure 8 shows that the Myc-tag could be completely removed by 5 μg/ml trypsin. The M13KO7 helper phage served as a negative control.

B. Generation of GM-UltraHelper phages:

215 Phage plaque assay was carried out using the supernatant containing GM-UltraHelper phage particles described above. A single phage plaque was picked and used to inoculate 10 ml 2x YT culture with 70μg/ml kanamycin. After 2 hours incubation at 37° C with constant shaking at 250 rpm, the culture transferred to a 2 liter

flask containing 500 ml 2x YT with 70 μg/ml kanamycin for large scale production of phage particles. The phages in the supernatants were then precipitated using polyethylene glycol (PEG)/NaCl, and re-suspended in phosphate-buffed saline (PBS). The phage concentration was determined by measuring OD_{268}. The OD OD_{268} measurement indicates that the culture contains approximately 2 X 10^{11}/ml GM-UltraHelper helper phage particles. Removing of GR2-Myc domains from the surface of phage by trypsin could increase the phage infectivity by 1 to 3 folds.

C. Use of GM-UltraHelper phage for displaying antigen-binding units:

216 In the subject adapter-directed display system, the exogenous polypeptide of interest is expressed as a fusion with an adapter (designated adapter 1) which interacts with a paring adapter (designated "adaptor2") that is fused in-frame with an outer-surface protein. The pairwise interaction between the two adapters facilitates display of the exogenous polypeptide. The phagemid vector pABMX14 is one of the expression vectors expressing an exogenous polypeptide fused in-frame with adapter 1. The vector pABMX14 (Figures 9A and 9B) was derived from pBluescript SK(+). A unique AgeI restriction site was introduced immediately after the lac promoter by PCR-based site-directed mutagenesis with a set of primers (pBS-Ska: 5'-GGAATTGTGAGCGGAT AACAATTTACCGGTCACACAGGAAACAGCTATGA-CCATG-3' and pBS-SKb: 5'-CATGGTCATAGCTGTTTCCTGTGTGACCGGTAAATT-GTTATCCGCTCACAAT T-CC-3'), and the XhoI and KpnI sites were deleted by cutting and blunt-end ligation. The synthetic DNA fragment flanked by AgeI at 5' and SalI sites at 3', containing ribosome-binding sequence RBS, pelB leader, and coding sequences for the adapter derived from $GABA_B$ receptor 1 (GR1, as adaptor1) and HA-$(His)_6$-tag (referred to as DH-tag), was cloned into the engineered pBluescript SK(+). The lac Z promoter drives

the expression of GR1 fusion and thus permits production of soluble exogenous polypeptide expressed with a bacterial cell.

217 To demonstrate that a functional protein can be displayed using the present invention, the single-chain antibody AM1 was subcloned into the pABMX14 vector. The resulting pABMX14-AM1 vector was transformed into TG1 cells, and the cells were superinfected with GM-UltraHelper phages with a multiplicity of infection (MOI) of 4 or 40 or 100. The phage particles were generated and purified as described in Example 1. The single chain antibody displayed on the phage surface was detected by phage ELISA using plates coated with AM1-antigen. The secondary antibody was HRP-conjugated anti-M13 antibody. The ELISA results showed that the phage particles generated from all three MOI infections exhibit binding specificity to the corresponding AM1-antigen, indicating that the single-chain antibody was functionally displayed on the phage surface (Figure 10). A dose-dependent binding was observed. Binding was saturated when the phage concentration reached 10^{12}/ml. The control phagemids generated from TG1 carrying pABMX14-AM1/ M13KO7 vectors did not bind to AM-1 antigen even at high concentrations such as 10^{13}/ml. The phage particles were also analyzed by Western blotting using anti-Myc and anti-HA antibodies. Phage particles were denatured by heating in SDS sample buffer under non-reducing condition (e.g. without β-mercaptoethanol). Figure 11 shows that the scFv antibody was only displayed upon infection with GM-UltrHelper phages and not with the control M13KO7 helper phages. The anti-Myc blot also revealed that approximately twice as much scFv-GR1-DH/GR2-Myc-pIII complex over the free GR2-Myc-pIII was assembled in phage particles in line 2. Since each phage particle contains 5 copies of pIII coat proteins. This indicates that each phage particles on average carry more than one copy of the scFv-GR1 fusion.

Example 3: Preparation of an Adapter-Directed Display System Comprising CM-UltraHelper phage vector

A. Construction of CM-UltraHelper phage Vector:

The pABMC13 vector was constructed by replacing the sequence between the XbaI and BglII sites of vector pABMD1 (Figure 22) with a synthetic DNA fragment comprising 5' to 3' a gene III leader sequence, a KpnI site, a coding sequence for Ala-Cys-Gly-Gly and a Myc-tag (Figure 12). This synthetic sequence was linked in-frame with gene III in pABMD1 vector.

The CM-UltraHelper phage vector was constructed by replacing the KpnI/BamHI fragment encoding (amino acid residues 11-19) and partial p III protein (amino acid residues 1-197) in the KO7Kpn helper vector with the corresponding fragment encoding a partial p III leader and the adaptor2-pIII fusion protein from the pABMC13 vector. The resulting CM-UltraHelper phage vector (Figure 13A) encodes an engineered pIII capsid fused with Cys-myc domain placed at the N-terminal of pIII. (Figure 13B). In addition, an amber stop codon TAG is placed between Cys-Myc coding sequence and gene III. Such a stop codon permits propagation of the entire phage particle in suppressor bacterial strains but not in non-suppressor strains.

The B9 KO7kpn helper phage clone was used for the construction of CM-UltraHelper phage vector. After subcloning of the engineered gene III fragment into KO7kpn phage vector, 24 kanamycin-resistant colonies were grown in 96-well microtiter plates in 2xYT medium with 70 ug/ml Kanamycin, and supernatants were used for phage screening by phage ELISA assay as described in Example 1. 23 clones were found to generate phage particles.

To confirm that CM-UltraHelper phages are able to package the expressed Cys-Myc-pIII fusions into the phage particles, ELISA assay using anti-Myc antibody was performed. The CM-UltraHelper phages from the selected five clones all displayed the Myc-tag on their surfaces. Such Myc-tag was not detectable in the KO7 helper phage negative control (Figure 14).

B. Generation of CM-UltraHelper phages:

218 Phage plaques formation assay was carried out using- the supernatant containing CM-UltraHelper phage particle according to the procedures described above. Briefly, a single phage plaque was picked and used to inoculate 10 ml 2x YT culture with 70μg/ml kanamycin. After 2 hours incubation at 37° C with constant shaking at 250 rpm, the culture was added to a 2 liter flask containing 500 ml 2x YT with 70 μg/ml kanamycin for incubation overnight. The phages in the TG1 supernatants were precipitated using polyethylene glycol (PEG)/NaCl, and re-suspended in phosphate-buffed saline (PBS). The phage concentration was determined by measuring OD_{268}. The phage yield for CM-UltraHelper helper phage is approximately 1-2 X 10^{12}/ml culture, which is very similar with that of M13KO7 and KO7kpn helper phage. The Myc-tag can be removed from the surface of phage by trypsin. Because of the amber stop codon placed in engineered gene III, no significant amount of CM-UltraHelper phage particles can be generated in a non-suppressor bacterial strain TOP10F'.

C. Use of CM-UltraHelper phage for displaying antigen-binding units:

219 The phagemid vector pABMX15 is another illustrative expression vector that expresses an exogenous polypeptide (a single-chain antibody AM1) fused in-frame with

an adapter. The vector pABMX15 (Figure 15A and 15B) was constructed from pABMD2 (Figure 22) by replacing the fd gene III fragment of pABMD2 (using NotI and SalI sites) with a synthetic DNA fragment encoding the HA-Tag and Gly-Gly-Cys.

220 To demonstrate that a functional protein can be displayed using CM-UltraHelper phage vector, the single-chain antibody AM1 was subcloned into the pABMX15 vector. The resulting PABMX15-AM1 was transformed into TG1 cells, and the cells were superinfected with CM-UltraHelper phage with a multiplicity of infection (MOI) of 1 or 10 or 50 or 100. Phage particles were generated and purified as described in Example 1. The single-chain antibody displayed on the phage surface was detected by phage ELISA using plates coated with AM1-antigen. 2×10^{12} phages were added for each well. The secondary antibody was HRP-conjugated anti-M13 antibody. The ELISA results showed that the phagemid particles generated from all four MOI infections were capable of specifically binding to AM1-antigen, indicating that functional single-chain antibody was displayed on the phage surface (Figure 16). The control phagemids generated from TG1 carrying pABMX15-AM1/ M13KO7 vectors did not bind to AM1-antigen. The phage particles were also used for western blot analysis. Phage particles were denatured by heating in SDS sample buffer under non-reducing condition (e.g. without β-mercaptoethanol). As shown in Figure 17, the scFv antibody was only displayed by CM-UltrHelper phage, but not M13KO7 helper phage.

221 Figure 17 also indicates that more free pIII is displayed on the phage particles than the AM1 scFv (see lane 4 of the left panel). This is indicative of monovalent display. By contrast, the GM-UltraHelper phage display system described in Example 2 yields more AM1 scFv than the free pIII (see lane 2 of Figure 11, left panel). The inclusion of the adapter sequence of $GABA_B$ receptor 1 in the scFv sequence, and the incorporation of the adapter sequence $GABA_B$ receptor 2 in GM-UltraHelper vector enhance the pairwise interaction.

Example 4: Preparation of an Adapter-Directed Display System Comprising GMCT-UltraHelper phage vector

A. Construction of GMCT-UltraHelper phage vector:

222 The pABMC12 vector was constructed from vector pABMC6, in which the Not1-BglII fragment was replaced with a synthetic DNA fragment containing coding sequence for Myc-tage, CT domain (amino acids 217-405) of gene III, a ribosome binding site (RBS) and an OmpA leader sequence (Figure 22A and B).

223 The GMCT-UltraHelper phage vector was constructed by replacing the KpnI/BamHI fragment in the KO7kpn helper vector with the corresponding fragment from pABMC12 vector. The resulting GMCT-UltraHelper phage vector (Figure 19A) encodes an additional copy of engineered pIII capsid, which comprises a GR2 domain, a myc-tag sequence (for detection of engineered pIII protein) and CT domain of pIII. Downstream to the engineered gene III, a ribosome binding sequence (RBS) and a leader sequence from the bacterial protein OmpA were fused to gene III sequences derived from pAMBD1. Those two copies of gene III-containing sequences are placed under the control of original gene III promoter (shown in Figure 19B). Phage ELISA assay was carried out to screen phage-positive clones as described in Example 1. 3 out off 10 clones were found to generate phage particles. Clone 3 was used for large-scale phage preparation.

B. Generation of GMCT-UltraHelper phage:

224 Phage plaques formation assay was carried out according to procedures described above. The phage yield for GM-UltraHelper helper phage was approximately 8 X 10^{11}/ml culture, which was similar with that of M13KO7 and KO7kpn helper phages. The GR2-Myc domains can be removed from the surface of phage by trypsin.

C. Uses of the GMCT-UltraHelper phage for expressing antigen-binding units:

225 The phagemid vector pABMX14 was used for phage display in combination with the GMCT UltrHelperelper phages. The single-chain antibody AM1 was subcloned into the pABMX14 vector (refer to as pABMX14-AM1). PABMX14-AM1 was transformed into TG1 cells, and the cells were superinfected with GMCT-UltraHelper phage with a multiplicity of infection (MOI) of 1 or 10 or 50 or 100. Phage particles were generated and purified as described in Example 1. The single chain antibody displayed on the phage surface was detected by phage ELISA using plates coated with AM1-antigen. The secondary antibody was HRP-conjugated anti-M13 antibody. The ELISA results demonstrated that the phagemid particles generated from all four MOI infections had similar activity binding to AM1-antigen, indicating that functional single-chain antibodies were displayed on the phage surface (Figure 20). The control phagemids generated from TG1 carrying pABMX14-AM1/ M13KO7 vectors exhibit no detectable binding affinity to AM1-antigen. The phage particles were also used for western blot analysis. As shown in Figure 21, the scFv antibody was displayed upon infection with GMCT-UltrHelper phages and not control M13KO7 helper phages.

Example 5: Enrichment of Phages Displaying Desired Polypeptides by Panning

226 A diverse DNA sequences can be cloned into either pABMX14 or pABMX15 vector for production of soluble polypeptides. This expression library can be used for displaying encoded polypeptide upon infection with the subject UltrHelper phages (GM and GMCT for pABMX14, CM for pABMX15). The specific protein or peptide displayed on phage can be enriched by several round of panning from a diverse library. The panning process is described as follows. Briefly, a 96-well plate is coated with specific antigens at a concentration of 1-10 ug/ml for overnight at 4°C. After wash with

PBS and blocking with 5% milk/PBS. 10^{11-12} phages are added and incubated for 2 hours at room temperature. After several washing with PBST and PBS, the bound phages are eluted with 10 ug/ml trypsin for 30 min, since all of subject UltraHelper phages have a cleavable Myc-tag fused to pIII protein. Trypsin elution is more efficient than 100 mM triethylamine (usually used in conventional phage panning) in our experiments. Upon repeating the process for several times, phages displaying the desired polypeptide can be enriched.

Express Mail Label No. EV339933445US

Example 6: Preparation and Uses of Bacterial Helper Vector

A. Construction of expression vector and bacterial helper vector:

227 The expression vector pABMX22 is constructed by replacing the sequence between the HindIII and SalI sites of vector pABMD1 (Figure 22A) with a synthetic DNA fragment encoding the GR1 adapter and HA-tag. As shown in Figure 25A, the vector contains an ampicillin-resistance gene for antibiotic selection (AMP), a plasmid replication origin (ColE1 ori), the f1 phage replication origin (f1 ori), and the lac promoter/lac O1 driving the expression of downstream sequence plac-RBS-p8L-GR1-HA-tag. The MluI/XbaI or MluI/NotI or XbaI/NotI restriction sites can be used to insert exogenous sequence for display or production of soluble protein in a bacterial cell. The complete vector sequence is shown in Figure 25B.

228 The bacterial helper vector pABMbd-1 (Figures 26A for vector map and 26B for the complete vector sequence) is derived from pBC-KS(+) vector from Stratagene. The sequences for multiple cloning sites between two BssHII sites in pBC-KS (+) are replaced by synthetic DNA fragment flanked by MluI sites (with a compatible cohesive end to BssHII) at 5' and BssHII sites at 3', containing ribosome-binding sequence RBS,

pelB leader sequence, the coding sequences for a chimeric outer membrane sequence consisting of the first 9 amino acids of E. coli major outer membrane lipoprotein (Lpp) and amino acids 46-159 of the outer membrane protein OpmA, and an adapter GR2 sequence. The lac Z promoter drives the expression of Lpp-OmpA-GR2 fusion, which will be secreted into the periplasmid of bacterial cells. The pABMbd-1 vector contains a chloramphenicol-resistance gene (Cam) for antibiotic selection, a ColE1 ori origin for plasmid replication, and the f1 phage replication origin (f1 ori) for phagemid package.

229

B. Generation of phagemid particle carrying bacterial helper vector:

230 The pABMbd-1 helper vector is transformed into bacteria TG1 cells. A single pABbd-1 colony is picked and used to inoculate 15 ml 2xYT culture with 50μg/ml chloramphenicol. After OD_{600} reach to 0.8, the bacterial cells are infected by KO7Kpn helper phage at MOI 10 for 1 hour at 37° C. The infected TG1 cells are cultured in a 2-liter flask containing 500 ml 2x YT with chloramphenicol and kanamycin for overnight. The phagemid particles in the supernatants were then precipitated using polyethylene glycol (PEG)/NaCl, and re-suspended in phosphate-buffed saline (PBS). The phagemid concentration is determined by measuring OD_{268}. The phagemid particles packaging pABMbd-1 helper vector are then used for adapter-directed bacterial display.

C. Use of bacterial expression and helper vectors for displaying antigen-binding units:

231 The scFv antibody gene AM2 is cloned into pABMX22 vector. The TG1 bacterial cells containing this expression vector is grown to OD_{600} = 0.8, and infected with the phagemid particle packaging bacterial helper vector pABMbd-1. The TG1 cells harbored expression vector and helper vector are grown in 2x YT with ampicillin/cam overnight at 30° C, then harvested and washed with PBS. Since the displayed proteins are tagged with HA-tag, the anti-HA tag antibody can be used for

detection the protein displayed on bacterial surface. For FACS analysis experiments, the cells resuspended in PBS are incubated with anti-HA tag antibody first, then incubated with fluorescein FITC-labeled anti-mouse antibody. After washing with PBS, the cells resuspended in PBS at 3-5 X 10^7/ml are counted on the basis of fluorescin intensity using a FACS sorter. The ELISA assay is also used for detect protein displayed on bacterial surface. First, AM2-antigen is coated onto ELISA plate for overnight at 4° C. After incubation with antigen for 2 hours, the cells bound to ELISA plate are detected with anti-HA antibody and HRP-conjugated anti-mouse antibody as described above.

232 A diverse antibody sequences can be cloned into pABMX22 vector for production of soluble antibody fragments. This expression library can be used to generate a selectable bacterial display library upon infection with the phagemid particle packaging the bacterial helper vector pABMbd-1. The bacterial cells displaying antibodies are incubated with the FITC-labeled antigen, and sorted on the basis of fluorescin intensity using a FACS sorter. After sorting, the selected cells are grown in 2XYT broth with AMP overnight. Subsequently, the cells are subcultured into fresh medium, and infected with bacterial helper vector for display. The displayed cells are then used to run through another round of FACS sorting selection.